SINGLE-MOLECULE BIOPHYSICS OF DNA CYCLIZATION

A Dissertation
Presented to
The Academic Faculty

By

Jiyoun Jeong

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Physics

Georgia Institute of Technology

May 2019

SINGLE-MOLECULE BIOPHYSICS OF DNA CYCLIZATION

Approved by:

Dr. Harold D. Kim, Advisor
School of Physics
Georgia Institute of Technology

Dr. James C. Gumbart
School of Physics
Georgia Institute of Technology

Dr. D. Zeb Rocklin
School of Physics
Georgia Institute of Technology

Dr. Nicholas V. Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Laura Finzi
Department of Physics
Emory University

Date Approved: February 28, 2019

But for me it is good to be near God;

I have made the Lord God my refuge,

that I may tell of all your works.

Psalm 73:28

To my love, Moka

ACKNOWLEDGEMENTS

I would like to thank Dr. Harold Kim for his generosity, patience, and support he showed me throughout my time as his student. I could not have been where I am without the opportunity he offered me to work in his lab and learn under his mentorship. I cherish every discussion I had with him.

I want to extend my thanks to my committee members, Dr. Laura Finzi, Dr. James Gumbart, Dr. Nicholas Hud, and Dr. Zeb Rocklin, for taking the time to read this dissertation and for providing valuable insights and advice. I especially thank Dr. Finzi for allowing me to work on a collaborative project and for inviting me to her lab.

I thank the former and present members of the Harold Kim Lab. I am grateful to Tung Le, James Waters, and Rasesh Parikh for sharing their knowledge and expertise during my early graduate years. I have been fortunate to be classmates, labmates, and friends with Gable Wadsworth and Bo Broadwater who have been supportive and understanding. I thank Derek Hart, Michael Ryan, and Alec Cook for their precious inputs and suggestions.

I want to thank my friends in Atlanta for their support and encouragement at various times. My special thanks to Micah and Jun for standing by my side when I needed.

I am indebted to the Hildebrand family who have been my family from the moment I first set foot in the United States. I am deeply grateful for their love and generosity.

I wish to thank my parents, brother, and sister for their constant love and support. I also want to thank my parents-in-law for their encouragement. Words cannot express the depth of my gratitude.

Finally, I thank Moka, my loving wife, for her endless love, forbearance, and sacrifice. Moka, you are my source of motivation and my best decision.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Structure and dynamics of DNA impact how the genetic code is processed and maintained. Thus, understanding the physical properties of DNA is of fundamental importance to biology. The physical properties of DNA have been extensively studied by a DNA cyclization assay, which measures the probability of intramolecular end-closure of DNA mediated by cohesive single-stranded overhangs called sticky ends. DNA cyclization kinetics are thought to have a clear theoretical basis, however, previous experimental studies produced conflicting results especially at lengths around 100 bp where DNA requires strong bending. The conflict partially arises from the incomplete understanding of how the intramolecular sticky-end joining reaction is influenced by the energetics and geometry of cyclized DNA. The main aim of this thesis is to bridge the gap between experimental measurables and theoretical predictions of DNA cyclization by investigating the association and dissociation kinetics of sticky ends of various DNA substrates. Using single-molecule fluorescence resonance energy transfer (FRET), we first show how the cyclization probability of short DNA depends on the rotational positioning of sticky ends around the helical axis. We find that rotational positions of sticky ends do not affect the cyclization rate, but allow the decyclization rate to oscillate, which we show to be related to the stackability of sticky ends at their termini. We also explore the effect of mismatches on DNA cyclization. Here, we present counter-intuitive findings that, for DNA cyclized by stackable sticky ends, base pair mismatch inside the DNA loop decreases the loop lifetime despite reducing the overall bending stress. This unexpected effect is contingent upon the terminal stackability of sticky ends and is most prominent for a central mismatch, furthest from the joined sticky ends. These findings show that base pair mismatch transfers bending stress to the antipodal side of the loop through an allosteric mechanism known as cooperative kinking. Based on our findings, we present a three-state model that correctly explains the apparent kinetics of DNA cyclization and decyclization.

CHAPTER 1

INTRODUCTION[1]

DNA is one of the most important molecules in life, as it contains all our genetic information. Consequently, its role as a genetic information repository has perhaps been the most emphasized aspect of the molecule. How the expression of this information is regulated, however, often depends on the mechanical properties of DNA that dictate its conformational changes, such as loop formation [1]. It has been well-demonstrated that DNA looping plays a mechanistically crucial role in gene expression pathways by modulating the binding and unbinding probabilities of gene-regulating proteins [2, 3, 4, 5, 6, 7]. In fact, the ability of DNA to bend and loop appears to be closely related to its underlying mechanisms and is not necessarily limited to regulation of gene expression, but is involved in many other cellular processes, including DNA packaging[8], replication [9, 10], and recombination [9, 10]. In this regard, the mechanical deformability of DNA is an equally important aspect of the molecule.

To elucidate the biophysical role of DNA as an elastic polymer, it is necessary to characterize the flexibility of DNA. Ideally, both the looping probability of DNA and the energetics of the looped configuration of DNA must be quantified as functions of DNA contour length. One experimental approach to studying the kinetics of DNA looping and unlooping is an assay known as DNA cyclization [11]. In this assay, DNA is prepared with cohesive single-stranded overhangs, called sticky ends, and mixed with ligase. In the presence of ligase, sticky-ended DNA molecules can either loop via intramolecular end-joining or dimerize (or multimerize) via intermolecular joining. Subsequently, the ligation kinetics of this reaction are used to quantify the looping probability, which is the key measurable of

[1]This chapter is partially excerpted from contents of a manuscript which is published in full as:
J. Jeong, T. T. Le, and H. D. Kim, "Single-molecule fluorescence studies on DNA looping". Methods. 2016;105:34-43.

this approach and reflects the intrinsic flexibility of the molecule [12]. Mechanical properties such as bending, twisting stiffness, and the helical period of DNA can then be deduced from the length-dependent looping probability [13] (a more detailed discussion is given in Section 1.3).

Despite its popular use and effectiveness, there exists a gap between experimental results and theory regarding DNA cyclization, particularly at short length scales near the smooth bending limit of DNA, which occurs around 100 bp [14, 15]. Consequently, both the applicability of this assay and the validity of experimental results at short length scales have been controversial [16, 17]. This controversy partially arises from the use of ligase[18] and the lack of understanding regarding the sticky-end joining reaction under the constraints imposed by such a small loop with high bending stress [16]. Although recently-developed ligation-free cyclization assays [15, 19] based on single-molecule fluorescence resonance transfer (FRET) appear to offer a promising approach that eliminates the concern of using ligase, the way in which the cyclization reaction is influenced by the joining and dissociation kinetics of sticky ends under the constraints imposed by such a small DNA loop remains elusive.

Motivated by this issue, this thesis examines the kinetic process of DNA cyclization at the single-molecule level, with the intention of bridging the gap between the experimental measure and theoretical prediction of short DNA cyclization. In particular, works included here will focus on elucidating the role of sticky ends in both DNA loop capturing and the breaking kinetics of DNA under strong bending conditions.

1.1 A note on terminology

Throughout this thesis certain terms are interchangeably used as described here. The terms, "overhang" and "sticky end" refer to an end of a double-stranded DNA at which several unpaired nucleotides of one strand extend beyond the other. The terms "annealing," "association," "duplex formation," "hybridization," and "joining" refer to formation

of double-stranded DNA from complementary single-stranded DNAs. The terms "cyclization," "looping," "loop capture," and "loop formation" refer to a process in which two ends of a double-stranded DNA are physically constrained in close proximity. The terms "cyclized DNA" and "looped DNA" refer to a conformational state of DNA in which the two ends are constrained in proximity by duplexed "sticky ends," or "overhangs." The terms "decyclization," "loop breakage," "loop release," and "unlooping" refer to a process in which the two constrained ends of a "cyclized DNA," or a "looped DNA" are allowed to move freely.

1.2 DNA structure and conformational flexibility

A brief introduction/reminder concerning DNA structure and how it is formed would be useful here. DNA is composed of molecules called nucleotides, each of which contains a five-carbon sugar (or deoxyribose), a phosphate group, and one of the following nitrogenous bases: adenine (A), thymine (T), guanine (G), or cytosine (C). A covalently-linked chain of nucleotides forms a single-stranded DNA (ssDNA) molecule. In ssDNA, the phosphate group at the $5'$ carbon site of the deoxyribose sugar is linked to the $3'$ carbon site of the sugar in the next nucleotide (see the structure colored in orange in Figure 1.1). These linkages between the phosphate and sugar molecules compose the backbone of the DNA strand with directionality. Under physiological conditions, DNA most stably exists as a double-stranded helical molecule. In double-stranded DNA, two oppositely-oriented ssDNA molecules are held together by pairing between complementary bases (A-T and G-C) as shown in Figure 1.1. In addition to base pairing, the structure of DNA is also maintained by an additional interaction known as stacking, which refers to attractive forces in aromatic–aromatic interactions between adjacent base pairs that are driven by electrostatic and hydrophobic effects. These interactions bring neighboring base pairs close together in a face-to-face alignment. Upon stacking, each base pair is slightly twisted in relation to its predecessor, by approximately $34°$. This angular shift makes DNA a right-hand helix in

solution, with a periodicity close to 10.5 bp per one helical turn. Individual contributions to the helical structure offered by pairing and stacking may be weak and easily broken by thermal energy. When many such interactions are present (e.g., for DNA longer than $\sim$30 bp), however, the structure of DNA at physiological temperature is quite stably maintained without disruptions in the helical structure. This structural stability can withstand mechanical deformation, such as bending and twisting, to some extent mainly because stacking gives base pairs some freedom to deviate from their preferred positions. The degree to which DNA resists mechanical deformation without permanent structural change characterizes how flexible it is as an elastic material.

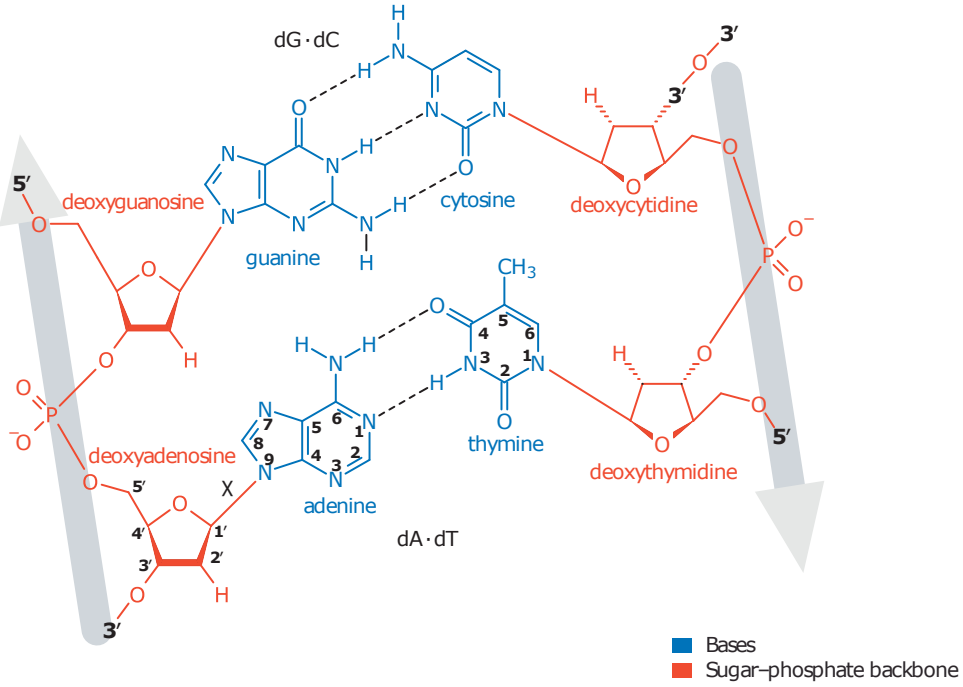


Figure 1.1: DNA chemical structure. Nucleobases are shown in blue and the sugar-phosphate backbones are shown in orange. Double-stranded DNA is formed by two complementary ssDNA paired in an anti-parallel manner. Grey arrows along the backbones point toward the 5′ end. The top and bottom base pairs represent G-C and A-T pairs, repectively. Dashed lines represent hydrogen bonding. Five carbon sites of deoxyribose are numbered in the structure of deoxyadenosine. This figure is redrawn from Figure 1 in reference [20] with permission from John Wiley and Sons.

1.3 Ligase-assisted DNA cyclization

Among various experimental approaches, ligase-based cyclization of DNA with sticky ends has been widely used to study the thermodynamics of DNA looping [13]. The most essential quantity to extract from the cyclization reaction is the cyclization probability which is known as the J factor (J). This quantity is measured in the units of concentration from the cyclization reaction as the ratio of the equilibrium constants for cyclization, K_C, and dimerization of a linear DNA molecule, K_D. Effectively, it reflects the probability of finding one end of a DNA chain in the vicinity of its other end. For extracting the J factor, the cyclization reaction does not directly measure K_C and K_D individually. Instead, it measures the ligation-dependent rate constants k_{cyc} and k_{dim}, which are the rates of circular monomer and linear dimer formation, respectively. The cyclization assay relies on ligase and gel electrophoresis. Ligase seals the discontinuous backbones, or nicks, near the region of joined sticky ends and converts the transiently-joined products into covalently-joined molecules such that they can be analyzed by gel electrophoresis. The ligated-products include monomer DNA circles as well as linear dimers or multimers, but they are easily distinguishable on a gel. By gel electrophoresis, the amount of monomer DNA circles is quantified as a function of time, from which k_{cyc} is extracted. For k_{dim}, the similar measurement can be done but using DNA with one sticky ends instead of two.

To connect the experimental measurables from the ligase-based cyclization assay (k_{cyc} and k_{dim}) to the J factor, consider the kinetic process of the cyclization reaction from the unlooped DNA molecule (L) to the final product of the closed DNA ring (P) as represented by the top schematic in Figure 1.2. The kinetic equation is given as the following [12]:

$$L \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} S + E \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} SE \xrightarrow{k_{34}} P + E$$

where S is a transiently cyclized molecule, E is ligase, and SE is the ligase bound cyclized DNA. Here, k_{12} and k_{21} correspond to the cyclization (or looping) and decyclization (or

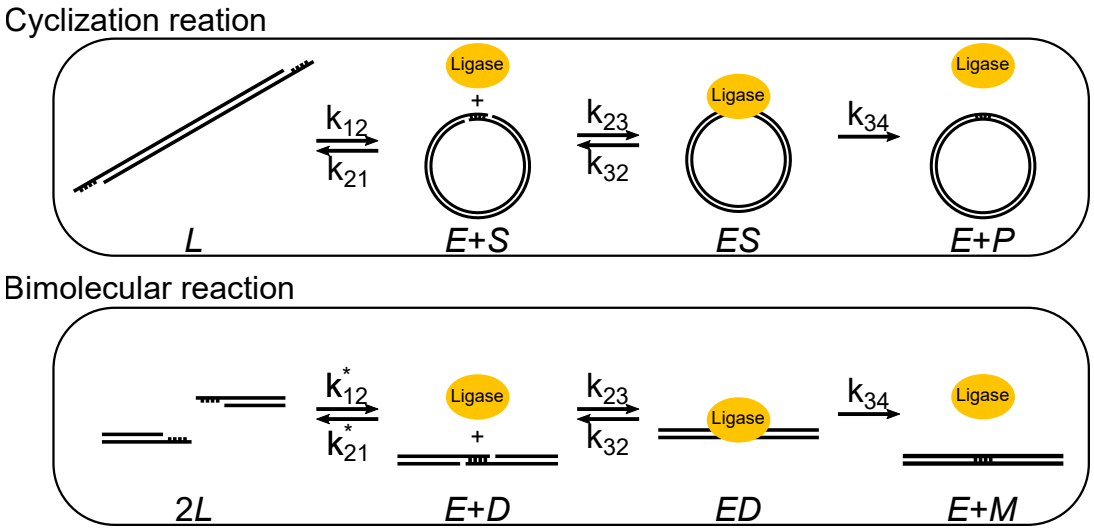


Figure 1.2: Schematics of ligation kinetics of DNA cyclization (top) and bimoecular joining reaction (bottom). Until ligase (E) is added, DNA molecules are in rapid equilibrium between the separated (L) and joined (S or D) states. Once ligase is added, the transiently joined molecules will be recognized by ligase at the rate of $k_{23}[E]$. The ratio between the accumulation rates of ligated products, P and M, will be equal to the cyclization probability.

unlooping) rates, and the ratio between these two is equal to K_C. k_{23} and k_{32} describe binding and unbinding rates of ligase onto the circular substrate, and k_{34} is the rate of covalent nick-closure. We can come up with a similar kinetic model for bimolecular joining with no looping involved (bottom, Figure 1.1), which is given by [12]:

$$2L \underset{k^*_{21}}{\overset{k^*_{12}}{\rightleftharpoons}} D + E \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} DE \xrightarrow{k_{34}} M + E$$

where D is a linear dimer and M is the linear ligated product. The forward and reverse rates of the bimolecular interaction between sticky ends are respectively denoted as k^*_{12} and k^*_{21}, with their ratio equal to K_D. The forward rates of producing P or M from L are equal to k_{cyc} and k_{dim}, respectively. k_{cyc} and k_{dim} are not only proportional to K_C and K_D, but also to k_{34} and ligase concentration, [E]. Thankfully, the ratio between these two can be a good approximation of the J factor, provided the following conditions are met:

1. The dissociation rate between sticky ends must be higher than the rate of forming the

ligatable substrate; that is:

$$k_{21} \gg k_{23}\,[E] \text{ (or } k_{21}^{*} \gg k_{23}\,[E] \text{ for dimerization).}$$

2. The fraction $(k_{32}+k_{34})/k_{23}$, known as the Michaelis constant (K_M), must be greater than $[S]$ (or $[D]$ for dimerization).

If these conditions are not satisfied, k_{cyc} and k_{dim} become insensitive to the concentration of ligase, thereby hiding the true equilibrium between the linear and circular, or the monomer and dimer fractions. Under conditions (1) and (2), while applying the steady-state approximation, k_{cyc} is given by

$$k_{cyc} \approx \frac{k_{34} K_C E_0}{K_M},$$

where E_0 is the total amount of ligase. Similarly, for dimerization,

$$k_{dim} \approx \frac{k_{34} K_D E_0}{K_M}.$$

Finally, taking the ratio between k_{cyc} and k_{dim}, we find the J factor, J, in the units of concentration; that is,

$$J = \frac{K_C}{K_D} \approx \frac{k_{cyc}}{k_{dim}}. \tag{1.1}$$

The concentration value given by the J factor is independent of the ligase concentration and the choice (i.e., sequence) of sticky ends.

Typically, the DNA cyclization assay is performed using DNA molecules of different lengths, as the length-dependence of the J factor can be used for validating polymer models of DNA. Here, it is also instructive to point out that ligation in the cyclization reaction forces the two ends of a cyclized molecule to be in a very specific orientation, in which the helical alignment is preserved across the joined ends. This implies that the length-dependence of the J factor would not only reflect the length-dependent bending elasticity, but also the torsional elasticity. Because the angular phase between the terminal base pair

changes as a function of DNA length, the ring-closure may require some degree of excess twisting depending on the length. For DNA lengths close to integer multiples of helical repeat (10.5 bp), excess twisting would not be necessary for cyclization; however, for lengths other than multiples of helical repeat, ligation certainly requires some extra twisting of the DNA molecule to align the loop termini. The consequence of this effect has been well demonstrated in many cyclization studies based on ligation as a strong oscillation in the J factor as a function of DNA length (Figure 1.3).

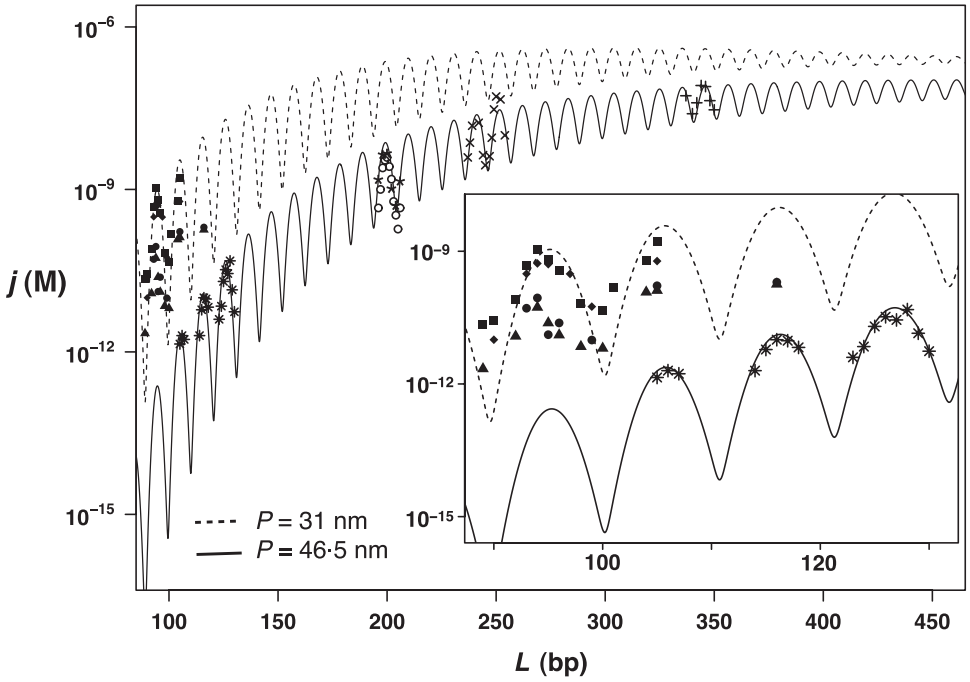


Figure 1.3: J factors from various cyclization studies are collected and plotted here as a function of DNA length, L. Experimental results are indicated by various symbols. Solid and dotted lines represent calculated J factors based on the WLC model with two different persistence lengths, P, indicated at the lower left corner and a helical repeat of 10.54 bp per turn. The J factor steeply decreases as L becomes smaller than $\sim$ 150 bp. The oscillation pattern in the J factor reflects the torsional constraints for ligation. Datasets are collected from the following studies: [21](times signs), [22](plus signs), [23](open circles), [14](black solid symbols), [18](asterisks), Rueter and Maher (unpublished results, stars). Coultier and Widom's data (black solid symbols near 100 bp (inset)) appear to be about 3 orders of magnitutde higher than the predication that fits the rest of the datasets. This figure is redrawn from Figure 1 in reference [13] with permission from Cambridge University Press.

1.4 The worm-like chain model

DNA locally exhibits a strong directional correlation, mostly due to the stacking interaction. This correlation is lost, however, if two distal base pairs along a long DNA strand, perhaps hundreds or thousands of base pairs apart, are considered. This "locally stiff, but globally flexible" property is one of the key characteristics of a semi-flexible chain, also known as a worm-like chain (WLC).

In the WLC model, a single parameter known as the persistence length (l_p) determines how strongly the directions between the two tangents along the chain are correlated. This directional correlation is lost according to $e^{(-L/l_p)}$, where L is the distance between the two tangents. Below its persistence length, a WLC would be generally linear, whereas at lengths well above this, the WLC would behave flexibly as a freely-jointed chain, where the directional correlation between each successive segment is completely lost.

The simplest way to represent DNA as a WLC is to coarse-grain DNA as a chain of short rods, each of which corresponds to one base pair. At this level of coarse-graining, there are only two angular degrees of freedom between adjacent links: θ and ϕ. θ is the difference in the tangent angle between two adjacent links, and ϕ is the dihedral angle or the azimuthal angle in a local reference frame. If there is no energy associated with these angular motions, the chain is a freely-jointed chain. If energy is associated with θ only, it is called a worm-like chain. If energy is associated with both angles, it is called a twisted worm-like chain. For WLC, the bending energy for each monomer is given as a quadratic function of θ, which is

$$E_{bend}(\theta) = \kappa\theta^2, \tag{1.2}$$

where κ is the stiffness constant. This stiffness coefficient is related to the persistence length. An equivalent expression of the bending energy in terms of the persistence length is

$$E_{bend}(\theta) = \frac{k_B T}{2}\left(\frac{L_p}{b}\right)\theta^2, \tag{1.3}$$

where k_B is the Boltzmann's constant, T is the temperature, and b is the monomer length of the chain. l_p has been measured using different methods in various conditions. The values can vary widely [24], but the consensus value for modeling purpose is around 50 nm [25].

1.5 J factor calculation

The two ends of a DNA molecule can approach each other due to thermal excitation, the probability of which is given by the Boltzmann factor. Without losing generality, we can fix one end of the chain at the origin to describe the extension of the chain with one position vector $\mathbf{r}$ of the free end of the chain. In equilibrium, the position vectors of the free end constitute the probability distribution or probability density function $P(\mathbf{r})$. Because the chain is inextensible, $P(\mathbf{r})$ is zero beyond one contour length ($L = Na$) away from the origin. Therefore $P(\mathbf{r})$ is normalized over the spherical volume of radius L:

$$\int_{|\mathbf{r}|\leq L} P(\mathbf{r}) d^3\mathbf{r} = 1. \tag{1.4}$$

For a free particle in the same spherical volume, $P(\mathbf{r})$ would be constant, equal to $1/\frac{4}{3}\pi L^3$. For a WLC like DNA, $P(\mathbf{r})$ depends only on the magnitude of $\mathbf{r}$. Hence, we can use the scalar end-to-end distance (r) to specify the macrostate of the chain. $P(r)$ can be calculated either analytically or numerically[26]. Examples of $P(r)$ are shown in Figure 1.4(A) for DNA of different lengths. The free energy of this macrostate ($A(r)$) is related to the radial probability distribution $P(r)4\pi r^2$:

$$\frac{e^{-A(r)/k_BT}}{Z} = P(r)4\pi r^2, \tag{1.5}$$

where Z is the normalization constant (or the total partition function). Therefore, $A(r)$ is given by

$$\frac{A(r)}{k_BT} = \log\Big(\frac{1}{4\pi r^2 Z P(r)}\Big). \tag{1.6}$$

This relationship is useful for depicting the free energy landscape of any transition associated with the change in the end-to-end distance r.

We can call a chain conformation as looped when the two ends are very close to each other. Mathematically, any conformation with the end-to-end distance less than some small threshold r_0 is a looped conformation. In the limit when $r \to 0$, $P(r)$ tends to a finite value. In chemistry and biology, it is more convenient to describe this probability density in terms of molarity. Converting the unit of $P(r \to 0)$ from number density to molar concentration give us a quantity known as the J factor.

$$J \equiv \lim_{r \to 0} P(r) \left[\frac{\text{moles}}{\text{L}} \right]. \tag{1.7}$$

Alternatively, the J factor can be thought of as the average concentration of one end in an infinitesimally small spherical volume around the other end,

$$J \equiv \lim_{r_0 \to 0} \frac{\int_{r=0}^{r_0} P(r) 4\pi r^2 dr}{\frac{4}{3}\pi r_0^3} \left[\frac{\text{moles}}{\text{L}} \right]. \tag{1.8}$$

The analytical formulae for the J factor of a homogeneous WLC were derived by Shimada and Yamakawa under different end-to-end constraints[27]. The result without angular constraints at the ends is plotted in Figure 1.4(B). The J factor is a nonmonotonic function of the chain length. At small lengths, the J factor is dominated by energy, whereas at large lengths, by entropy. Thus, the J factor peaks near 500 bp (top, Figure 1.4(B)), which indicates that sites that are 500-bp apart have the highest probability of juxtaposition. The J factor changes steeply at lengths below 100 bp, which is better appreciated on a semilog plot (bottom, Figure 1.4(B)). The absolute J factor values and the peak position slightly change when juxtaposition of two interior points of the chain is considered, but the overall trend remains similar[28].

To include anisotropic, asymmetric, or nonuniform flexibility and curvature, it is necessary to obtain the J factor by simulation. The goal of the simulation is to generate a set

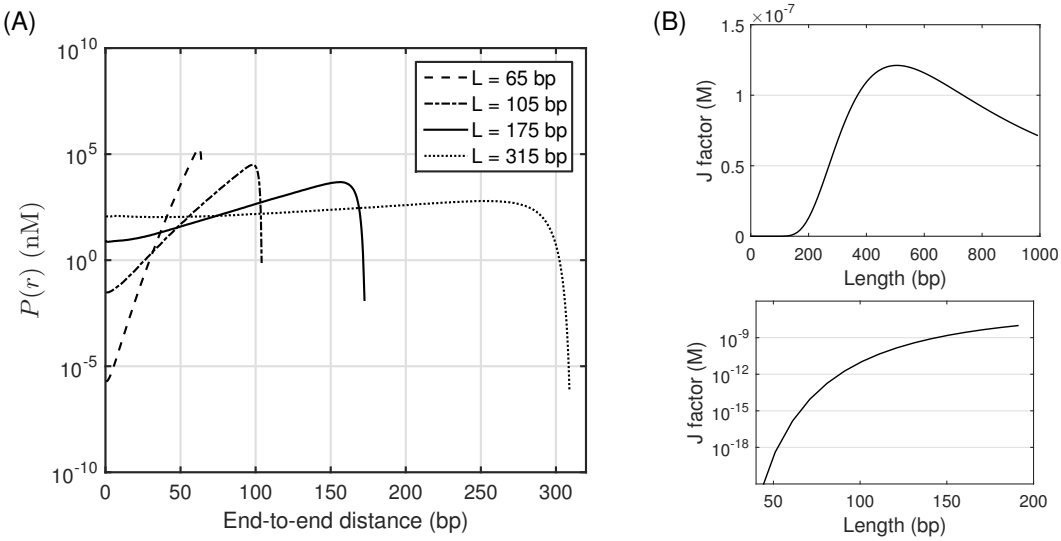


Figure 1.4: Probability density function and J factor. (A) Probability density function $P(r)$ for different lengths of DNA. These functions are obtained by the simulation method described in 1.5. Because DNA is modeled as an inextensible chain, $P(r)$ drops off near the contour length of the chain. As the length becomes shorter, $P(r)$ approaches the y-axis at a steeper angle. Where these curves cross the y-axis are equal to the J factors. (B) J factor vs. length. These are calculated using the formula[29] originally derived by Shimada and Yamakawa[27] using the persistence length of $50\,\mathrm{nm}$. The top plot shows the J factor on a linear axis over a wide range of DNA lengths. The bottom plot is focused on the controversial region of the relationship on a log y-axis.

of chain conformations in thermal equilibrium, which is the canonical ensemble in thermodynamics. The simplest method to perform this task is the Gaussian sampling method[30, 31], which exploits the normal distribution of bending angles dictated by the Boltzmann distribution of energy (Eq. 1.3). One typically generates $10^6 - 10^7$ chains and builds a histogram of end-to-end distances ($N[(i-1)\Delta r \leq r < i\Delta r]$) using equally spaced bins (Δr). Normalizing this histogram by the total number of conformations yields the probability corresponding to $4\pi r^2 P(r)\Delta r$. Dividing this by the thin shell volume $4\pi r^2 \Delta r$, one can obtain the probability density function $P(r)$. The J factor can be read off from the y-intercept of $P(r)$ (Figure 1.4(A)).

To compute $P(r)$ of the J factor of DNA shorter than about one persistence length, this free sampling method does not yield a statistically significant number of looped conformations in the ensemble of $10^6 - 10^7$ conformations. Therefore, one has to perform a biased

sampling method called the umbrella sampling. In this method, one tries to restrain the two ends of the chain at close range by linking them with a stiff spring. In this method, chain conformations cannot be generated by the Gaussian sampling method because of the restraint on the end-to-end distance. Instead, one has to perform a Monte Carlo (MC) simulation where random perturbations of θ are followed by the Metropolis-Hastings acceptance criterion. From a set of MC simulations with different spring potentials (U_j), a set of biased histograms centered around different end-to-end distances can be constructed. The stiffness of the spring (K) and undeformed spring lengths ($r_{0,j}$) have to be chosen carefully to cover the entire range of the end-to-end distance while allowing the histograms to be overlapped significantly. Once a full set of histograms are constructed, the unbiased probability distribution can be found using the weighted histogram analysis method (WHAM)[26, 32, 33]. Specifically, the unnormalized radial probability density in the i-th bin (p_i) is given by

$$\begin{aligned} p_i &= \frac{\sum_{j=1}^{N_{sims}} n_{i,j}}{\sum_{j=1}^{N_{sims}} n_{i,j} f_j c_{i,j}}, \\ f_j = \frac{1}{\sum_{i=1}^{N_{bins}} c_{i,j} p_i}, &\quad c_{i,j} = e^{-U_j/k_B T}, \quad U_j = \frac{1}{2} K (r - r_{\circ,j})^2, \end{aligned} \tag{1.9}$$

where N_{sims} is the number of simulations, N_{bins} is the number of bins used to construct the histograms, $n_{i,j}$ is the number of conformations in the i-th bin ($N[(i-1)\Delta r \leq r < i\Delta r]$) of the j-th simulation, and $U_j(r_i)$ is the bias factor evaluated at r_i in the j-th simulation. To find the optimal set of $\{f_j\}$, the above equations are calculated iteratively, assuming an initial set of $\{f_j\} = 1$, and updating the equations until $\{p_i\}$ converge. Finally, dividing the entire set of $\{p_i\}$ by $4\pi r^2 \Delta r$ gives $P(r)$. This calculation procedure is summarized in Table 1.1 using a 150-bp-long DNA chain as an example.

1.6 Short DNA cyclization

Recently, there has been interest in finding experimental evidence of the WLC prediction of the J factor at very short length scales, well below the persistence length of DNA. Several

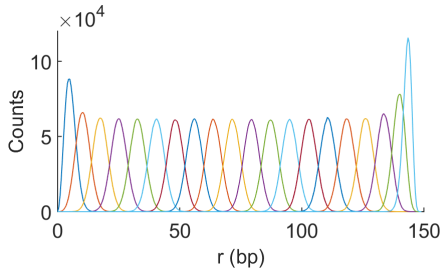
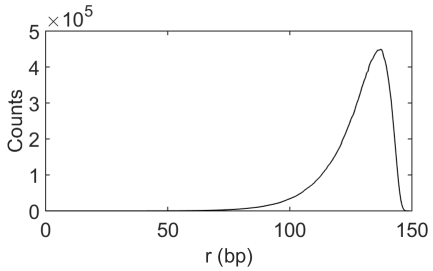
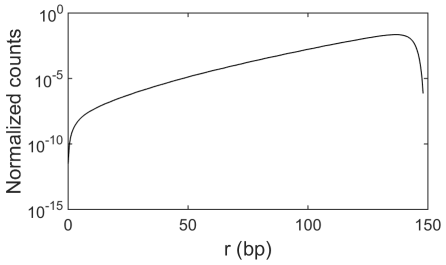
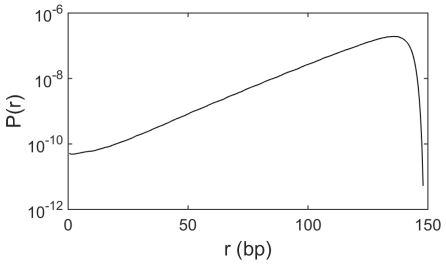
Quantity	Computational procedure	Corresponding plots ($L = 150$ bp)
$n_{i,j}$	Perform the umbrella sampling to obtain histograms of end-to-end distances with the bin-width of Δr ($N_{sims} = 20$).	
p_i	Perform WHAM to recover the unbiased histogram of end-to-end distances.	
$\frac{p_i}{\sum p_i}$	Divide by the total number of conformations.	
$\frac{p_i}{\sum p_i 4\pi r^2 \Delta r}$	Normalizing the histogram amplitude by $4\pi r^2 \Delta r$ yields the probability density function ($P(r)$).	

Table 1.1: Step-by-step instructions with intermediate plots

cyclization studies have tested the WLC model of DNA at short length scales [14, 18, 34, 35], but the results of these studies have been controversial. Cloutier and Widom repeatedly observed the J factor of DNA near 100 bp to be several orders of magnitude greater than the WLC prediction (see the inset in Figure 1.1)[14, 35]. Vologodskii and coworkers, however, argued that the results of these experiments were biased due to high ligase concentration [18]. Under the correct ligation conditions, the J factor appears to be valid down to a contour length of approximately 100 bp [18, 34]. Nonetheless, the possibility of higher-than-WLC flexibility of short DNA has been raised theoretically by including the effect of spontaneous structural changes in DNA that increase local flexibility, such as bubble or kink formations [36, 37]. Because cyclized short DNA is under strong conformational stress, the influence of bubble or kink formations appears plausible, although experimental evidence supporting enhanced flexibility due to these formations is lacking. It still seems valid, however, to consider that cyclization under strong bending conditions may be dramatically different from that under negligible bending stress, such as in the case of dimerization or cyclization of a long DNA. Additionally, the kinetics of sticky-end joining and dissociation under strong cyclization conditions must also be carefully considered. It has been suggested that the nicks near the joined sticky ends are highly prone to kink formation with small cyclized DNA [16]. Furthermore, the dissociation rate of joined sticky ends is a sensitive function of the unlooping force when the loop size is small.

1.7 DNA cyclization at the single-molecule level

It would be quite difficult to address concerns regarding short DNA cyclization, especially regarding the kinetics of sticky ends, with the conventional ligase-based cyclization assay because it cannot separately measure the looping and unlooping rates (k_{12} and k_{21}), or the joining and dissociation rates (k_{12}^{*} and k_{21}^{*}) of sticky ends. Additionally, the use of ligase (see Section 1.3) can further limit the feasibility of the method, especially at the short length scale where cyclization is rare.

A ligase-free, single-molecule approach, recently developed independently by the Ha lab [15] and the Kim lab [19], provides a feasible alternative to the ligase-based method. This newer method has several advantages over the ligation method, one of which being that it does not rely on the ligation reaction, allowing the cyclized and decyclized states to be separately monitored. The Ha group measured the J factor in the range of 67 to 200 bp, which is the controversial length range. Remarkably, in this study, the J factor in this range did not show a strong length dependence, unlike the prediction from the WLC model. In addition to the deviation from the WLC prediction, the length-dependent oscillation of the J factor, which was attributed to the helical nature of DNA, is also puzzling given that sticky-end joining does not require the precise helical alignment across the nicks.

Our lab has also developed a single-molecule DNA cyclization assay [19], similar to the method of the Ha grouop. Our protocol, however, differs from theirs, mostly notably in DNA sample preparation, in which we employ polymerase chain reaction (PCR) and gel-purification to minimize sequence errors. In the following chapters, I will explain our method in greater detail and present our work on short DNA cyclization, using various DNA substrates and sticky-end designs that could shed light on the issues introduced above.

1.8 Structure of the thesis

In Chapter 2, I will provide detailed experimental protocols for the FRET-based DNA cyclization and further discuss how the experimental measurable from the FRET-based assay is related to the theoretical definition of the J factor. In Chapter 3, I will explore how sticky ends influence the looping and unlooping rates of short DNA using our FRET-based assay. In Chapter 4, I will investigate the role of base pair mismatch in DNA cyclization and present how the sharp kink at the mismatch influences the sticky-end joining and dissociation reactions. Finally, in the last chapter, I will present some preliminary data from the on-going research and discuss possible future experiments in relation to our findings that could further our understanding of DNA mechanics.

CHAPTER 2

SINGLE-MOLECULE FRET STUDY OF DNA CYCLIZATION[1]

2.1 Introduction

Although the ligase-based DNA cyclization assay successfully validated the worm-like chain (WLC) model prediction with long DNA [13], it has serious limitations when it comes to looping of short DNA [14, 18]. First, because cyclization is rare for short DNA, the ligation reaction is dominated by dimerization, and the interpretation of the result becomes challenging. It was shown that high ligase concentration can bias the apparent dimerization constant [18], thus causing an overestimation of the J factor at short length scales. Another drawback of the assay is the ligase itself, which nonspecifically binds DNA, loses activity over time during reaction, and only works in a limited range of salt conditions and temperature.

Recently, researchers began to use single-molecule fluorescence resonance energy transfer (smFRET) to tackle questions on DNA bendability [15, 33, 54, 57]. By immobilizing DNA molecules with 7–10 base long sticky ends on a glass surface, one can monitor reversible looping and unlooping events in real time under a microscope without competing dimerization as in bulk cyclization [54]. Moreover, because end-to-end annealed DNA is torsionally relaxed around the nicks, the looping equilibrium is not influenced by torsional rigidity of DNA. Unlike the irreversible ligase-dependent cyclization assay, the smFRET assay also gives access to the unlooping transition, which can give us insights into the energetics of sharply bent double-stranded (dsDNA) [33]. In this chapter, we introduce a FRET-based single-molecule cyclization assay that our lab previously used to study ther-

[1]This chapter is partially excerpted from contents of a manuscript which is published in full as:
J. Jeong, T. T. Le, and H. D. Kim, "Single-molecule fluorescence studies on DNA looping". Methods. 2016;105:34-43.

modynamics of DNA looping and unlooping.

2.2 Experimental setup

2.2.1 FRET

FRET refers to a radiationless resonance energy transfer mechanism by which energy from a donor fluorophore in the excited state is transferred to an acceptor fluorophore in the ground state via long-range dipole–dipole interaction. Forster showed [58] that the rate of the energy transfer, k_{FRET}, from the excited-state donor to the ground-state acceptor depends on the distance between the two fluorophores, r, and the quantity R_0, known as the Förster distance. In his definition, k_{FRET} is given by

$$k_{FRET} = \frac{1}{\tau_D}\left(\frac{R_0}{r}\right)^6 \tag{2.1}$$

where τ_D is the fluorescence lifetime of the excited donor in the absence of the acceptor. The expression of R_0 is given by

$$R_0 = \left(\frac{9000(\ln 10)\kappa^2 Q_D I}{128\pi^5 n^4 N_A}\right)^{-6} \tag{2.2}$$

where Q_D is the fluorescence quantum yield of the donor in the absence of the acceptor, n is the index of refraction, κ^2 is the dipole orientation factor (which is equal to 2/3 for a freely-rotating dipole pair), N_A is Avogadro's number, and I is the spectral overlap integral between the donor emission and the acceptor excitation. FRET, therefore, requires the donor and acceptor pair to be in close proximity, and the emission spectrum of the donor and the excitation spectrum of the acceptor to be overlapped.

The fraction of the donor emission used for FRET defines the FRET efficiency, E_{FRET}, which is given by

$$E_{FRET} = \frac{k_{FRET}}{\tau_D^{-1} + k_{FRET}} = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \tag{2.3}$$

Practically, using a fluorescence microscope, E_{FRET} is approximately found by the relative proximity ratio (E_{raw}) between the donor and acceptor intensities, which is given by

$$E_{raw} = \frac{I_A}{I_D + I_A} \tag{2.4}$$

where I_D and I_A are the background-subtracted donor and acceptor intensities, respectively. This ratiometric FRET efficiency differs from E_{FRET} due to factors such as leakage of the donor emission to the measured acceptor intensity and differences in quantum yield and detection efficiency between the donor and the acceptor. To find E_{FRET} from the measured intensities, E_{raw} can be corrected as [98]

$$E_{FRET} = \frac{I_A - \beta I_D}{\gamma I_D + (I_A - \beta I_D)} \tag{2.5}$$

where β is a correction factor for the donor leakage and γ is a correction factor for the differences in quantum yield and detection efficiency between the donor and the acceptor, respectively. β can be measured by using donor-only substrates. γ can be found from the change in the donor and acceptor intensities (ΔI_D and ΔI_A) before and after acceptor photobleaching as $\gamma = \Delta I_A / \Delta I_D$ [99]. At $r = R_0$, k_{FRET} is equal to τ_D^{-1} and $E_{FRET} = 0.5$. Thus R_0, which depends on the spectral properties of the donor and acceptor pair, defines the fluorophore-specific characteristics of FRET.

In our experiments, we use Cy3 and Cy5 fluorophores as the donor and acceptor FRET pair. These dye molecules can be covalently attached to various positions on a DNA molecule such as the 5′ end, the 3′ end, a nucleobase, or internally to the sugar-phosphate backbone in place of a base. Such labeling is commercially available. For the Cy3 and Cy5 FRET pair, R_0 is found to be approximately 6 nm [59].

2.2.2 Total internal reflection fluorescence microscopy

Total internal reflection fluorescence microscopy (TIRFM) is a widely-used optical imaging technique for detecting fluorescent molecules near the surface with a high signal-to-noise ratio to the low background intensity. In TIRFM, the excitation is achieved by a laser beam that is totally internally reflected at the interface of the surface of a flow chamber and the aqueous buffer inside. The main advantage of the beam with TIR is that it creates an exponentially decaying evanescent wave in a direction perpendicular to the specimen surface, which can selectively illuminate fluorescent samples near the surface. For this reason, TIRFM is typically used for illuminating samples that can be immobilized on a surface.

To achieve the total internal reflection (TIR) excitation, the excitation beam must strike the glass surface at an angle, θ_i, greater than the critical angle, θ_c, as defined by Snell's law, which can be expressed as,

$$\theta_i > \theta_c = \arcsin\left(n_{buffer}/n_{glass}\right) \tag{2.6}$$

where n_{glass} and n_{buffer} are the refractive indices of the specimen surface and the aqueous buffer, respectively. For $n_{glass} \approx 1.5$ and n_{buffer} equal to that of water, $\theta_c \approx 62°$.

2.2.3 Microscope setup

We achieve TIR excitation via an oil-immersion objective lens. In this objective-based TIRF microscope setup, the excitation laser beam is aligned to enter the objective lens from the bottom side near the edge of the objective aperture, such that the excitation ray can come out angled (Figure 2.1(A)). Additionally, for the angled incident ray to be properly collimated, the laser beam must be focused at the back focal place of the objective lens. In our setup, we use a small broadband mirror mounted on a translation stage below the objective lens to introduce the excitation beam into the objective lens from the bottom. Unlike many other TIR setups where a dichroic mirror is used for this purpose, we prefer

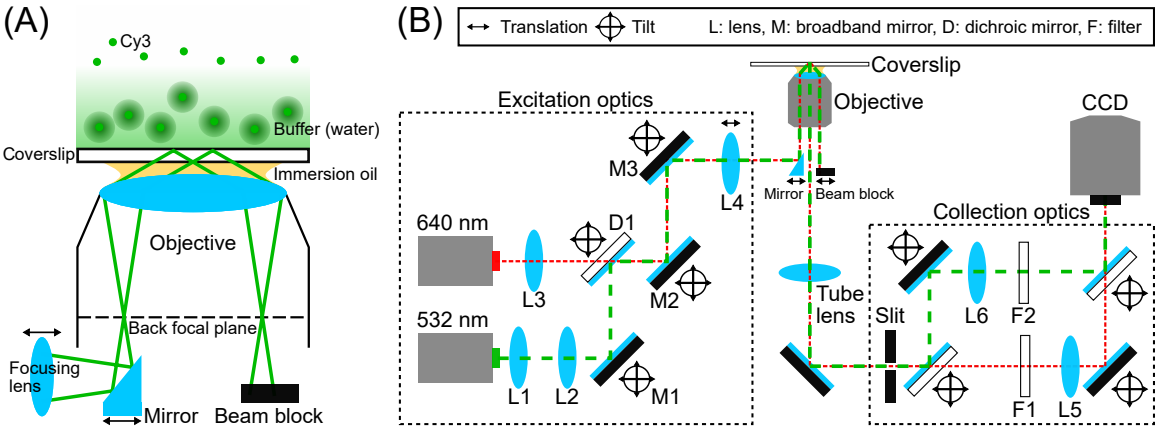


Figure 2.1: Objective-type TIRFM. (A) Principle of objective-type TIR excitation. The excitation beam passes through the focusing lens and strikes the small broadband mirror below the objective lens. Both the focusing lens and the mirror are on translation stages to have the beam focused at the back focal plane, close to the edge of the objective lens. As a result, the exiting beam is steeply angled and totally internally reflected at the buffer-coverslip interface, generating an evanescent field in side a flow chamber in the direction normal to the surface. The reflected beam entering back into the objective is blocked by a beam block. (B) TIRFM setup. In the excitation optics, the 532 nm laser is expanded via L1 and collimated by L2. Since the 640 nm beam comes out from a fiber optic cable, it does not need to be expanded, but is collimated by L3. Using M1 and D1, we make the 532 nm beam to be on the same path as the 640 nm beam as much as possible. By M2 and M3, both 532 nm and 640 nm beams are align to hit the small mirror below the objective lens to acheive TIR excitation. M2 and M3 are also used in pair to ensure TIR excitation is the strongest near the center of the objective's field of view. The emission signal from the sample comes out to the side-port of the microscope body and is cropped by a mechanical slit. The cropped image is splitted by a dichroic mirror into two, which are further filtered by F1 for Cy5 emission and F2 for Cy3 emission. By L5 and L6, the magnification of the separated images are adjusted to be similar to each other in magnification. Finally, the filtered images are recombined side-by-side at the CCD imaging port.

using the small broadband mirror since it allows multi-color excitation, and provides a better emission signal by minimizing filtering [143]. The translation stage of the mirror allows the beam to be near the edge of the objective lens. The angle of the beam coming out of the objective (θ_i) is determined by the numerical aperture (NA) of the objective lens. Objectives with high NA greater than 1.45 are generally required for practical reasons [60], and the objective lens (UApo N 100X/1.49 Oil TIRF) used in our lab sufficiently satisfies this requirement.

In Figure 2.1(B), Our excitation and collection optics setup is illustrated. To excite the

fluorophores (Cy3 and Cy5) in our experiments, we use two separate solid-state lasers at 532 nm and 640 nm. The 532 nm laser (NT66-968, B&W Tek) is used to excite the donor and the 640 laser (CUBE 640-30FP, Coherent) is used for direct excitation of the acceptor, which is required to confirm the co-existence of the acceptor molecule at locations where donor-labeled samples are found. For collection optics, we employ a two-color collection setup in which the fluorescence emission from Cy3 and Cy5 is separated by the wavelength via a dichroic mirror (FF545/650-Di01-2536, Semrock). The separated emission beams are then filtered and projected side-by-side onto an EMCCD (DU-897ECS0-# BV, Andor).

2.2.4 Temperature control

We maintain the temperature of the specimen by controlling the temperate of the objective lens via an objective collar (Bi-150303, Automate Scientific) through which thermalized water from a refrigerated water bath/pump (Isotemp 1006s, Fisher) can circulate. The temperature of the specimen at the contact point with the thermalized objective lens is measured and calibrated by a thin thermocouple probe (TJC36-CASS-010G-6-SMP-M, Omega) placed inside a flow chamber.

2.2.5 Surface preparation for sample immobilization

Because TIFRM is specialized for surface-based experiments, the surface of the sample chamber must be cleaned, passivated, and functionalized for specific tethering of fluorescent molecules. Here, I describe surface preparation methods for surface-based fluorescent experiments.

First, microscope slides and coverslips are cleaned using the following protocol:

1. If starting with a fresh microscope slide, drill eight pairs of holes along the two long edges of the slide to make inlets and outlets using a 0.75-mm diamond drill bit. Each pair of drilled holes at the opposite edges serves as a single flow channel.

If reusing a slide, soak the slide in water overnight and boil it in water for 10 min in a microwave oven. Remove any visible residues from previously-used epoxy or double-sided tape.

2. Rinse drilled slides and coverslips with deionized (DI) water and sonicate them in water for 10 min.

3. Dry slides and coverslips in a vacuum chamber.

4. Place slides and coverslips in an upright position in a plasma cleaner/etcher (PDC32G, Harrick Plasma) and etch them for 10 min at the HIGH RF level.

At this stage, the surfaces of slides and coverslips become hydrophilic and ready for passivation. The surfaces can be passivated by either PEGylation or the dimethyldichlorosilane (DDS) and Tween-20 protocol [95].

PEGylation

The PEGylation protocol is as follows:

1. Mix 80 mg of mPEG-silane (MW = 2000) and 2 mg of biotin-PEG-silane (MW = 3400) in 320 μL of 100 mM sodium bicarbonate solution.

2. Pipette ∼40 μL of the PEG mixture onto a clean slide and place a coverslip over the wetted slide such that the PEG solution is spread evenly across the space between the slide and coverslip.

3. After 1 hour of incubation in the dark, separate the slide and coverslip, rinse them with DI water, and dry with compressed air. The above PEGylation protocol skips the amine modification of the surface [61] by using mPEG-silane.

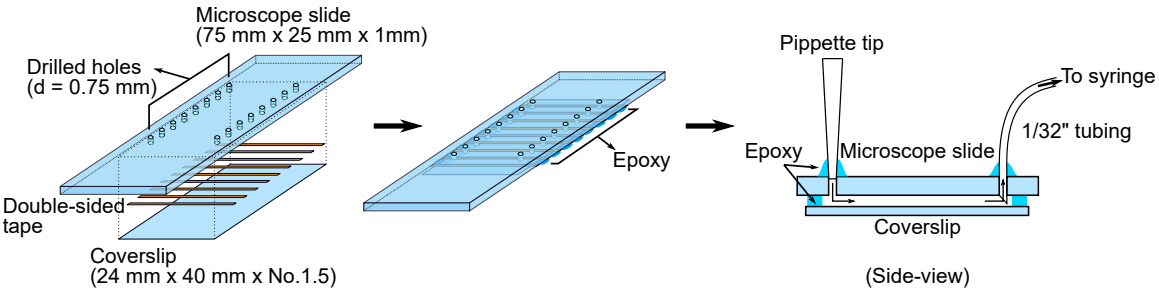


Figure 2.2: Flow chambers are made by putting a microscope slide and a glass coverslip together with double-sided tape and sealing with epoxy. Buffer and sample can be injected into the flow cell directly using a pipette. Alternatively, a motorized syringe pump can be used. For this setup, PEEK tubing can be additionally attached to either holes of a flow cell. The other end of the tubing is connected to the syringe pump, which withdraws buffer and sample inside a pipette tip into the flow cell.

DDS-Tween-20 passivation[*]

Alternatively, DDS-Tween-20 passivation can be employed, as follows:

1. Rinse slides and coverslips twice with spectroscopy-grade hexane.

2. Place slides and coverslips in a slide-staining container pre-cleaned with hexane by sonication. Fill the container with hexane to minimize air inside the container and add 50 µL of DDS. Tightly seal the container and shake for 1.5 hr.

3. After incubation, rinse slides and coverslips with hexane and then sonicate in hexane for 2 min.

4. Repeat Step 3 three times.

[*]This protocol must be performed under a fume hood.

After the passivation step, a flow chamber is assembled by placing a slide and a coverslip with double-sided tape and epoxy as illustrated in Figure 2.2.

For sample immobilization, we employ a strong binding interaction between NeutrAvidin and biotin. For the PEGylated flow channel, 50 µL of NeutrAvidin (100 µg/mL) prepared in T50 (10 mM Tris-HCl, 50 mM NaCl, pH 8.0) is flowed in and incubated in the

channel for approximately 3 min. After the incubation, NeutrAvidin is washed with T50. At this point, the flow channel is ready for tethering biotinylated DNA molecules to the surface.

For the DDS-Tween-20 flow channel, 50 μL of biotinylated bovine serum albumin in T50 buffer (0.2 mg/mL) is flowed in. After 10 minutes, the flow channel is washed with T50, followed by flowing in 100 μL of 0.2% Tween-20 in T50, which is incubated in the channel for 10 minutes. After the Tween-20 incubation, the NeutrAvidin solution and biotinylated DNA molecules are introduced into the channel, as with the PEGylated flow channel.

2.2.6 DNA preparation

There are two different ways to construct DNA molecules for the DNA cyclization assay. One way is to synthesize or order DNA oligonucleotides that are fluorescently labeled near the ends with donor and acceptor fluorophores. For surface immobilization, biotin also needs to be incorporated either near the end or near the center of the DNA. One should note that immobilization of DNA to the surface near its center could lead to a slightly higher looping probability [62]. A more efficient and accurate method to make DNA molecules for this assay is to use PCR (see Figure 2.3) [19]). In this method, one orders two pairs of PCR primers that contain the required modifications (Cy3, Cy5, and biotin) and uses them to amplify the source DNA (plasmid or genomic DNA) in separate PCR reactions. The DNA products are heated and cooled, and as a result of strand exchange, DNA carrying the sticky ends (7–10 bp) and all three modifications can be unambiguously identified on the surface. The big advantage of this method over the first one is the ease of making long DNA molecules and a much lower rate of introducing sequence errors [19].

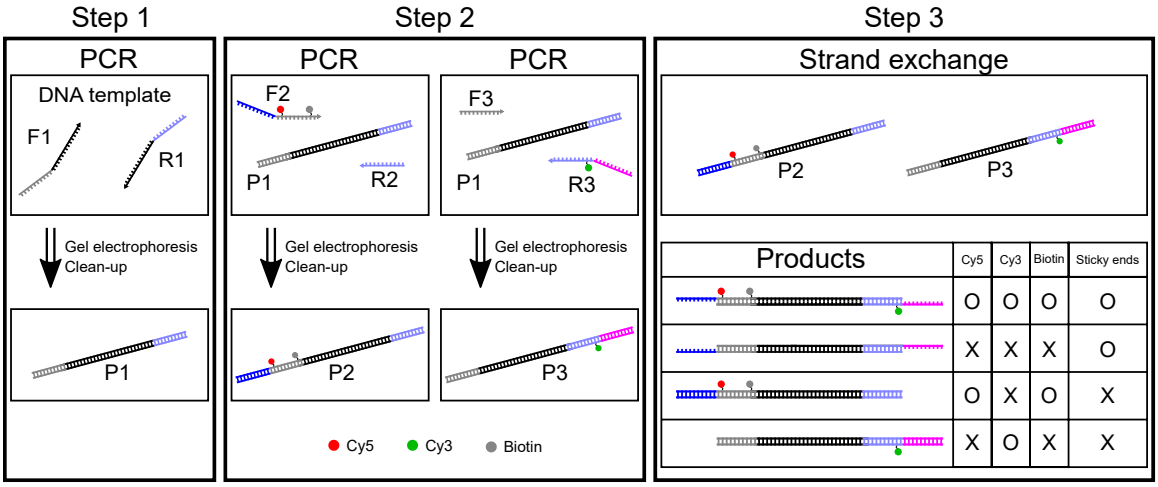


Figure 2.3: Sticky-ended DNA with modifications, Cy3, Cy5, and biotin, is prepared in three steps. In Step 1 (*left*), we extract sample DNA of interest from the source DNA (DNA template). During the PCR reaction, the two ends of the sample DNA is replaced by common sequences present in PCR primers (colored segments in F1 and R1), which we call adapters, for the ease of modification later. The PCR product (P1) is then purified by gel electrophoresis and a PCR clean-up kit. For storage and sequencing, P1 could be incorporated (cloned) into a plasmid DNA. In Step 2 (*middle*), P1 (or preferably the plasmid containing P1) is separately PCR-amplified with two primer pairs, F2 and R2, and F3 and R3. The F2 and R2 primer pair incorporate Cy5, biotin, and the left-side sticky-end sequence (shown as blue in F2) into P1. The other primer pair, F3 and R3, incorporate Cy3 and the right-side sticky-end sequence (shown as magenta in R3), which is complementary to the left-side sticky-end sequence. All four primers contain the adapter sequences. The PCR products from these PCR reactions, which are P2 and P3, are separately cleaned and purified by gel and a PCR clean-up kit. Finally in Step 3 (*right*), we mix P2 and P3 into one tube with a ratio of 1:4 (P2:P3) in a buffer containing 10 mM Tris-HCl, 1 mM EDTA, and 150 mM NaCl. The mixture is heated to 95° and slowly cooled to 4° over the course of 2 hrs. The strand-exchange reaction between P2 and P3 can result in four different molecules, which are summarized in the table in Step 3. In principle, the 1:4 ratio generates about ∼80% of the molecule with all modifications (i.e., Cy5, Cy3, biotin, and sticky ends). In practice, we find that ∼65 to 70% of the product with all modification.

2.2.7 Data processing and analysis

In-house Win32 Visual C++ software is used for operating an EMCCD, the 532 nm and 640 nm lasers, and a motorized syringe pump. The same software is also used for recording a series of 2×2 binned images captured by the EMCCD camera. The raw image data are stored as a 3D array of 16-bit binary integers (i.e. image width × image height × frames), with each element representing a raw intensity value.

The raw data are processed by in-house Matlab GUI code to locate co-localized spots with Cy3 and Cy5, and to generate intensity trajectories. Before image processing, the Matlab code requires a set of calibration points to map the Cy3 channel onto the Cy5 channel. Such calibration points are manually found by the user by locating spots that show an anti-correlating intensity switch indicative of FRET. Once the control points are specified, the mapping between the two channels can be simply done using the "fitgeotrans" function in Matlab, with the "transformationType" input argument specified by either "affine" or "lwm".

To locate DNA molecules with fluorescence, a background-subtracted image averaged over the last ten frames is first constructed. Based on this image, high intensity pixels on the Cy3 channel are selected by thresholding. The corresponding Cy5 pixels are then automatically found by the mapping function.

For each pair of selected spots from the two channels, the local background around the peak is subtracted from the intensity values calculated for 3×3 pixels to generate an intensity time trajectory. The corresponding FRET trajectory is then calculated based on the donor and acceptor intensity values from the time trajectory.

2.2.8 smFRET looping assay

To monitor the cyclization reaction using FRET, sticky-ended DNA molecules with modifications (i.e. Cy3, Cy5, biotin) are sparsely immobilized onto the surface of a flow channel. Typically, the DNA solution is diluted at $\sim$50 pM and about 20 µL to 50 µL is flowed in a time, until desired density ($\sim$100 to 200 molecules per field of view) is achieved. Once the molecules are immobilized, the flow channel is washed with T50 to remove excess, untethered DNA molecules. To prevent the immobilized fluorescence molecules from photobleaching and blinking, we use an imaging buffer containing the PCD-PCA oxygen scavenging system[107] in Trolox. The oxygen scavenging system in the buffer removes oxygen, which react with the fluorophores and make them photo-intactive. Trolox prevents

photoblinking of the fluorophores by quenching triplet states. Once the flow channel is filled with the imaging buffer, the molecules are excited by the 532 nm laser. The emission in the green channel indicate the DNA molecules in the unlooped state, while the red channel shows the molecules in the looped state. The 640 nm laser is briefly turned on together with the 532 nm laser at the start and end of each recording to directly excite the acceptor molecules and to co-localize the DNA molecules with active donor and acceptor pairs.

One can perform this smFRET assay perturbatively by changing the salt concentration of the imaging buffer (similar to stopped-flow experiments) or non-perturbatively in one salt concentration. The non-perturbative method allows measurements of both looping and unlooping rates, but may not yield good statistics for both transitions depending on DNA length and salt concentration. In this case, a salt jump experiment is preferred [15, 33].

2.3 Experimental measurables from the FRET looping and unlooping assays

Although the WLC model has proven to be a robust coarse-grained model of DNA, it recently faced some challenges, especially from studies on short DNA looping. The controversy is not so much as to whether the WLC model fails or not, but rather, when and why it fails because, after all, Eq. 1.2 is only an approximation for small bending angles. The first in a series of studies came from the Widom group. They measured the J factor of DNA at or below 100-bp length using the ligase-dependent cyclization assay. The measured J factor deviated from the theoretical prediction by several orders of magnitude[14]. Some theories were proposed to explain this phenomenon, most notably, by invoking local melting[36, 39]. A year later, the same cyclization method was used to measure the looping probability at similar lengths by the Vologodskii group, but the result upheld the WLC model[18]. Recently, the DNA looping problem was investigated using single-molecule fluorescence[15, 54]. Especially, the Ha group performed the J factor measurement with short DNA and extracted J factor values seemingly much higher than the WLC model prediction. This study reignited interest of the field on this topic. However, technical concerns

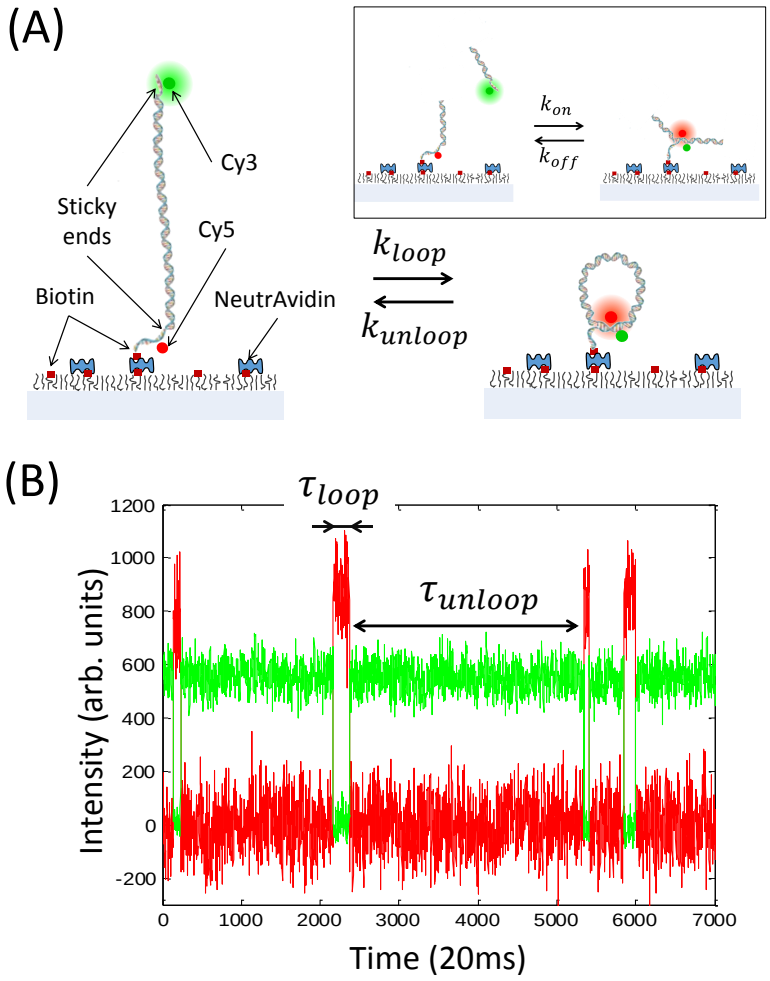


Figure 2.4: Single-molecule FRET looping assay. (A) The experimental setup. Double-stranded DNA molecules with complementary overhangs (sticky ends) and the FRET pair (Cy3 and Cy5) are immobilized on the glass coverslip, and fluorescence signals are observed by an objective-type total internal microscope. DNA molecules can be trapped in the looped state due to base pairing between the sticky ends. They loop and unloop with apparent rates of k_{loop} and k_{unloop}, respectively. The boxed inset shows the single-molecule measurement of the association (k_{on}) and dissociation (k_{off}) rates between the unlinked sticky ends. (B) Time trajectories of Cy3 and Cy5 fluorescence intensities. The intensities fluctuate between two levels due to reversible looping and unlooping. The FRET efficiency is low in the unlooped state and high in the looped state. Individual dwell times in the looped (τ_{loop}) and unlooped (τ_{unloop}) states are used to build dwell time histograms, from which the unlooping rate (k_{unloop}) and the looping rate (k_{loop}) can be determined.

about the extraction of the J factor[16] and salt effects[51] remain.

The measurables in the smFRET assay (Figure 2.4(A)) are the looping rate (k_{loop}), the unlooping rate (k_{unloop}), or the ratio of the two (k_{loop}/k_{unloop}). In the smFRET assay, loop-

ing and unlooping of DNA lead to fluorescence intensity fluctuations. High Cy3 and low Cy5 signals correspond to the unlooped state whereas low Cy3 and high Cy5 signals correspond to the looped state (Figure 2.4(B)). From the intensity fluctuations, individual dwell times in the looped state (τ_{loop}) and the unlooped state (τ_{unloop}) are recorded and averaged to obtain the mean dwell times or lifetimes ($\langle\tau_{loop}\rangle$ and $\langle\tau_{unloop}\rangle$). The looping and unlooping rates are equal to the inverses of these lifetimes: $k_{loop} = \langle\tau_{unloop}\rangle^{-1}$ and $k_{unloop} = \langle\tau_{loop}\rangle^{-1}$. These apparent rates, however, depend on the affinity between the sticky ends. Therefore, to extract quantities such as the J factor attributable to DNA conformation only, the looping and unlooping rates must be normalized against the rates for association (k_{on}) and dissociation (k_{off}) between the sticky ends. These rates are measured in a similar FRET experiment using the truncated end segments of the DNA molecule (inner panel, Figure 2.4(A)).

In this section, we carefully examine the relationship between these measurables and the J factor and contend that the J factor cannot be unambiguously determined from these measurables. Therefore, although the apparent rates provide valuable insights into DNA bendability, their relationship to the J factor must be interpreted with caution.

2.3.1 Looping rate

First, let's consider a normal bimolecular reaction between two complementary single-stranded DNA (ssDNA) molecules. Without losing generality, we consider one to be fixed in space and the other one freely diffusing about. Association (annealing) between them can be schematized by a pseudo first order reaction:

$$\mathrm{A} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathrm{AB} \tag{2.7}$$

where A is the single-stranded state of A, and AB is the duplex state where base pairs are formed with B. The association rate of B to A (k_{on}) is proportional to the concentration of

B ($[B]$), and therefore

$$k_{on} = k'_{on}[B]. \tag{2.8}$$

The second-order rate constant k'_{on} can be measured from the concentration dependence of k_{on}. The dissociation rate of B from A (k_{off}) is concentration independent.

In the smFRET assay, the high FRET state is stabilized by base pairing between the sticky ends of a DNA molecule. Association between two ends of the same molecule can be similarly described using a first-order reaction

$$\text{A} \underset{k_{unloop}}{\overset{k_{loop}}{\rightleftharpoons}} \text{AB} \tag{2.9}$$

where k_{loop} and k_{unloop} are the first-order rates for low-to-high and high-to-low FRET transitions, respectively. k_{loop} should be proportional to the effective concentration of one end of the DNA molecule in close proximity to the other end, which is the J factor (J). Hence, we can simply use J as the concentration in Eq. 2.8 to relate J to the measured k_{loop},

$$k_{loop} \approx k'_{on} J \qquad \text{or} \qquad J \approx \frac{k_{loop}}{k'_{on}} \tag{2.10}$$

Therefore by dividing the first-order rate of loop formation (k_{loop}) by the second-order rate constant of duplex formation between the sticky ends (k'_{on}), one can extract the J factor in molar units. This is the expression used in our previous study[54]. In comparison, Vafabakhsh and Ha divided $k_{loop} + k_{unloop}$ by k'_{on} to extract the J factor in their study[15]. This expression, however, is expected to overestimate the J factor as k_{unloop} increases as DNA loop becomes shorter (explained in Eq. 2.11 and Figure 2.5(A)). Implicit in Eq. 2.10 is the assumption that k'_{on} is not affected by the presence of the loop.

2.3.2 Unlooping rate

Experimentally, looping of DNA shorter than one persistence length is extremely rare, which motivates experimenters to use extreme conditions (high salt, high ligase concentration). We find that a more robust way to test the WLC model in standard salt conditions is to measure the unlooping rate of a small DNA loop. In this assay, DNA molecules with sticky ends are first prepared in the looped state by using a high salt buffer, which are then induced to break open by switching to a low salt buffer. From the exponential decay of the looped population, the unlooping rate (k_{unloop}) or the looped-state lifetime ($\langle \tau_{loop} \rangle$) can be extracted. This loop lifetime decreases as the loop size becomes smaller because of higher bending stress.

This relationship can be understood based on the transition state theory. Imagine the transition state of duplex dissociation is Δr away from the ground state at r_0. r_0 is the length of the duplex between the sticky ends in the ground state. According to the transition state theory, the dissociation rate is proportional to the probability of occupying the transition state. The shear force by the loop lowers the free energy of the transition state more than that of the ground state by ΔA (Figure 2.5). Therefore, k_{unloop}, which is the dissociation rate of the duplex holding the loop, would be faster than that without the loop (k_{off}, Eq. 2.7) by a Boltzmann factor:

$$k_{unloop} = k_{off} \exp(\Delta A / k_B T). \tag{2.11}$$

Using a linear approximation, the free energy difference can be expressed in terms of the shear force (f) exerted on the duplex formed between the sticky ends of the loop (Figure 2.5(A)):

$$k_{unloop} = k_{off} \exp(f \Delta r / k_B T), \tag{2.12}$$

which is equivalent to the well-known Bell model of force-induced bond breakage[53]. f

can be calculated as a function of loop size L by differentiating the free energy by end-to-end distance. But this requires knowing the functional form of $P(r)$ at small r, which is complicated even in approximate forms[26]. Instead, one can run an umbrella sampling simulation with a restraining potential around r_0[33]. The force can then be easily computed from the variance of the end-to-end distances[52]. Using this unlooping assay, we previously showed that the linear relationship between the logarithm of the loop lifetime and f breaks down at the loop size of 60 bp with sodium only and 100 bp with 5 mM magnesium[33]. The mechanism behind this critical limit still needs to be investigated. Based on previous experimental[50, 51] and computational[40, 41, 42] studies, it likely reflects the kinking transition of a dinucleotide. Fraying around preexisting nicks could also relax the bending stress[44, 45], but such transition seems to be preceded by kinking[49].

2.3.3 Equilibrium fraction

Another method to extract the J factor experimentally is to measure the equilibrium fraction of the looped state. This method has been used in the analysis of tethered particle motion experiments[3, 46, 47], and appears to be applicable to smFRET experiments that record looping and unlooping in a reversible manner[54]. From single-molecule trajectories of looping and unlooping, one can calculate the ratio of the looped time (τ_{loop}) to the unlooped time (τ_{unloop}). Separately, one can measure the affinity between the molecules that stabilize the DNA loop (protein-protein, protein-dsDNA, or ssDNA-ssDNA interaction) (K_D), similar to Eq. 2.7. The equilibrium constant K_D in molar units is measured by dividing the first order dissociation rate constant (k_{off}) by the second order association constant (k'_{on}). Using a simple thermodynamic argument with the assumption that K_D is the same in the context of the loop, the J factor can be extracted from three measurables as

$$J \approx K_D \frac{\tau_{loop}}{\tau_{unloop}}. \tag{2.13}$$

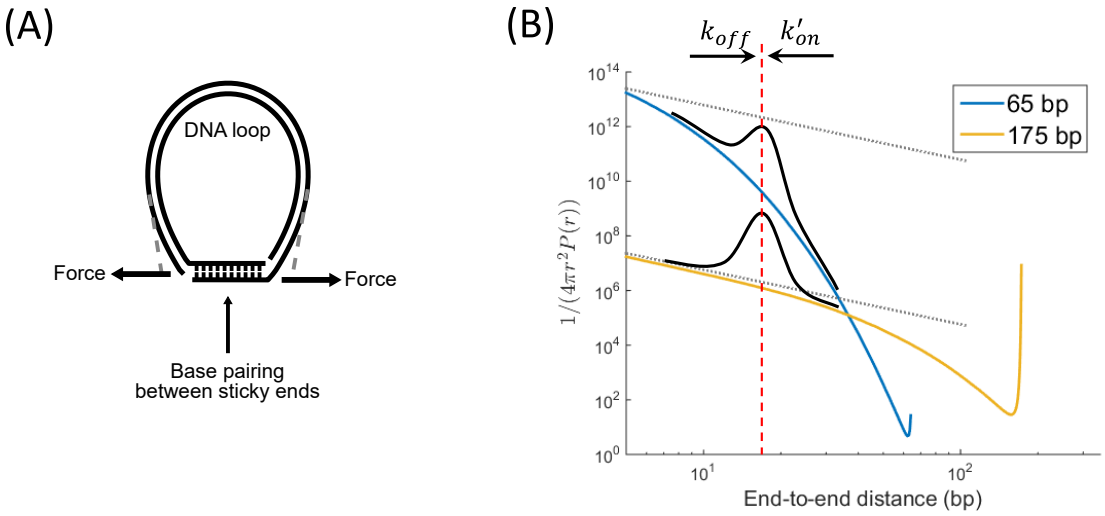


Figure 2.5: The influence of a tightly bent DNA on base pairing between sticky ends. (A) Schematic of a DNA loop stabilized by end-to-end annealing. The duplex formed between the sticky ends experiences a shear force due to the elasticity of the DNA loop. (B) The free energy landscape of duplex association and dissociation. The blue (65 bp) and yellow (175 bp) curves are the inverse probability density functions used in Eq. 1.6, which are calculated from the simulation as described in 1.5. On a log y-axis, the inverse probability density function can be interpreted as the free energy landscape (Eq. 1.6). The x-axis is also scaled logarithmically to highlight the slope of the landscape. The black curves represent the transition barrier that the end-to-end distance must cross for duplex formation or dissociation. The red dashed line indicates the transition state position. Short DNA (blue) has a heavily tilted landscape which increases duplex dissociation rate (k_{off}) and decreases duplex association rate (k'_{on}). For reference, the landscapes corresponding to the association of unlinked sticky ends is shown as gray dotted lines.

Expressing this in terms of rates, we obtain

$$J = \frac{k_{off}}{k'_{on}} \frac{k_{loop}}{k_{unloop}} < \frac{k_{loop}}{k'_{on}}, \tag{2.14}$$

where the latter inequality holds because $k_{unloop} > k_{off}$ according to Eq. 2.11. Therefore, the J factor extracted this way is smaller than the J factor extracted from Eq. 2.10. The discrepancy between the two expressions will grow larger as the DNA loop becomes smaller.

The takeaway message is that the dependence of the J factor on measurable rates and equilibrium constants is more complicated than generally assumed. The J factor was origi-

nally formulated for long DNA[11, 48], where the effect of the loop on k'_{on}, k_{off}, and K_D would be negligible. The effect of the loop on base pairing equilibrium can be visualized by the free energy landscape. In Figure 2.5, the free energy landscape of the end-to-end distance r of DNA is plotted for 65-bp (blue) and 175-bp (yellow) DNA. For comparison, the free energy landscapes from contant $P(r)$'s are plotted as gray dotted lines, which follow a power law with exponent equal to -2. Overlaid in black are the transition barriers for duplex formation. For long DNA (175 bp), the free energy landscape at small r has a similar slope to the gray line, which indicates that base pairing equilibrium between sticky ends would not be greatly affected by the loop. In contrast, the free energy landscape for the short DNA(65 bp) is heavily tilted compared to the gray line, which lowers the transition barrier for k_{off} and raises it for k'_{on}. Therefore, the assumption of loop-independent k'_{on} used in Eq. 2.10 or loop-independent K_D used in Eq. 2.13 is no longer valid for short dsDNA.

A more theoretically correct treatment of the J factor is given only recently by Mulligan et al.[2, 43] by solving the Fokker-Planck equation using the minimum free energy path. In this model, both the looping and unlooping rates scale with the J factor. Especially, the looping rate (k_{loop}) scales with the J factor with an exponent less than one, whose values depends on the reaction distance and geometry.

2.4 Conclusion

In conclusion, we introduce single-molecule fluorescence assays to study biophysics of DNA loops. The enzyme-free smFRET looping assay overcomes many of the technical challenges presented by the ligase-dependent cyclization assay. Since the smFRET assay can also be performed in a reversible manner, one can focus on the unlooping transition, which is well suited to studying the energetics of strong DNA bending. Although this smFRET assay and other looping assays produce rates that give us insights into elastic properties of DNA, we argue that extraction of the J factor from these rates is highly prob-

lematic, especially for small DNA loops where the loop stabilizing interaction is under high shear force.

CHAPTER 3

DETERMINANTS OF CYCLIZATION-DECYCLIZATION KINETICS OF SHORT DNA WITH STICKY ENDS[1]

3.1 Introduction

DNA under physiological conditions constantly undergoes conformational changes due to thermal fluctuations. Among those changes, bending motions coupled with twist can bring distal sites into proximity [117] and impact genome packaging and gene regulation [113, 115]. Some of these processes involve looped DNA segments much shorter than 500 bp, a length regime where the bending energy begins to dominate the free energy of loop formation. For example, some operons in *Escherichia coli*, such as *lac* and *gal*, are regulated by repressor proteins that form loops as small as $\sim$100 bp [112]. Small DNA loops can also be induced by some restriction endonucleases [91, 96, 118] or actively extruded by chromosome packaging motor proteins [125]. In many cases, proteins stabilize small DNA loops that spontaneously arise; therefore, it is of great importance to quantify the probability of spontaneous looping events. On the other hand, the protein complexes that bridge two distal sites of short DNA segments are subjected to a significant amount of bending and torsional stress depending on the loop geometry and size [43, 71]. This stress can affect the binding affinity of the protein complexes, and thereby alter the lifetime of the looped state [2, 91, 114]. Recently, small DNA loops have also been used as force sensors and applicators to study bending mechanics of DNA itself or force-dependent conformational changes of other biomolecules [51, 80, 84, 85, 94]. Therefore, measuring looping and unlooping dynamics of short DNA segments can give us insights into the energetics and

[1]This chapter is based on a manuscript in preparation for submission as:
J. Jeong and H. D. Kim, "Determinants of cyclization-decyclization kinetics of short DNA with sticky ends".

internal forces that govern loop-associated processes and applications.

The simplest way to form DNA loops is to use DNA with two complementary single-stranded overhangs, or sticky ends, in a reaction called cyclization. In this reaction, the sticky ends of the same DNA molecule hybridize to each other to form a "linker" duplex. To a good approximation, the cyclization (looping) rate (k_{loop}) is thought to be the product of two quantities [12]: (i) the effective concentration of one sticky end in the proximity of the other, which is known as the J factor (J), and (ii) the annealing rate constant between the two sticky ends (k_{on}). Therefore, if k_{on} is known, the J factor can be determined by measuring k_{loop}. The J factor can also be predicted from polymer models as a function of length, deformability, and loop geometry. Hence, the J factor has been used as a hallmark to test and refine DNA models such as the worm-like chain model.

Nonetheless, the experimental attempts to measure the J factor of DNA shorter than one persistence length ($\sim$150 bp) have so far been controversial. Using a ligation-based assay, the Widom group first measured the J factor of short DNA molecules [14]. This study reported an anomalously high J factor, but the anomaly was soon proven to be an artifact due to the high concentration of ligase in a study by Vologodskii [18]. In a more recent study, Vafabakhsh and Ha used a ligase-free fluorescence resonance energy transfer (FRET) assay to measure the J factor of short DNA molecules in the range between 50 and 200 bp [15]. The reported J factor displayed an oscillatory pattern as a function of DNA length, which indicated that the apparent cyclization kinetics depend on the torsional degree of freedom. However, the DNA length-dependent oscillatory pattern from this FRET-based cyclization study remains puzzling because it is out of phase with that from the ligation-based cyclization study [16, 17].

To shed light on this unresolved issue, we investigate how DNA cyclization and decyclization rates are influenced by the torsional degree of freedom: the rotational positioning of the sticky ends around the helical axis and base stacking between the sticky ends. Using the single-molecule FRET assay, we measured both cyclization (looping) and decycliza-

tion (unlooping) rates of short DNA ($\sim$ 100 bp) over two helical periods with either (1) full sticky ends that allow terminal base stacking or (2) gapped sticky ends that prevent terminal base stacking. We find that the cyclization rate varies monotonically with DNA length for both sticky-end types, whereas the decyclization rate shows length-dependent oscillation only with full sticky ends, fast at half-integer number of helical turns and slow at integer number of helical turns. Based on separately measured dissociation kinetics of sticky-ended duplexes, we attribute this kinetic difference between integer and half-integer loops to both the terminal base stacking and the shearing geometry imposed by the loop. The J factors extracted from our measured cyclization and bimolecular hybridization rates are in agreement with the worm-like chain prediction down to $\sim$90 bp despite uncertainties due to sequence and experimental condition. We also explain the origin of the oscillatory J factor reported in the previous study [15] and discuss inherent uncertainties in the experimentally derived J factor that may hamper an accurate comparison to theory, especially for DNA shorter than 100 bp.

3.2 Materials and methods

3.2.1 Preparation of DNA molecules with sticky ends

We obtained two different master DNA molecules from phage lambda DNA and yeast genomic DNA by polymerase chain reactions (PCR). We performed a second set of PCR reactions on these master templates to produce DNA set 1 and set 2 whose lengths range from 96 bp to 116 bp and 108 bp to 124 bp, respectively (see Appendix A). By PCR primer design, both sets of DNA shared common 20-bp "adaptor" sequences at the ends. In each set, a DNA molecule was lengthened by inserting base pairs immediately before the two adaptor regions.

To make DNA molecules amenable to the surface-based FRET looping assay, we performed additional PCR reactions on unmodified DNA molecules with primers carrying the adaptor sequences and the necessary modifications (i.e. FRET donor, FRET acceptor,

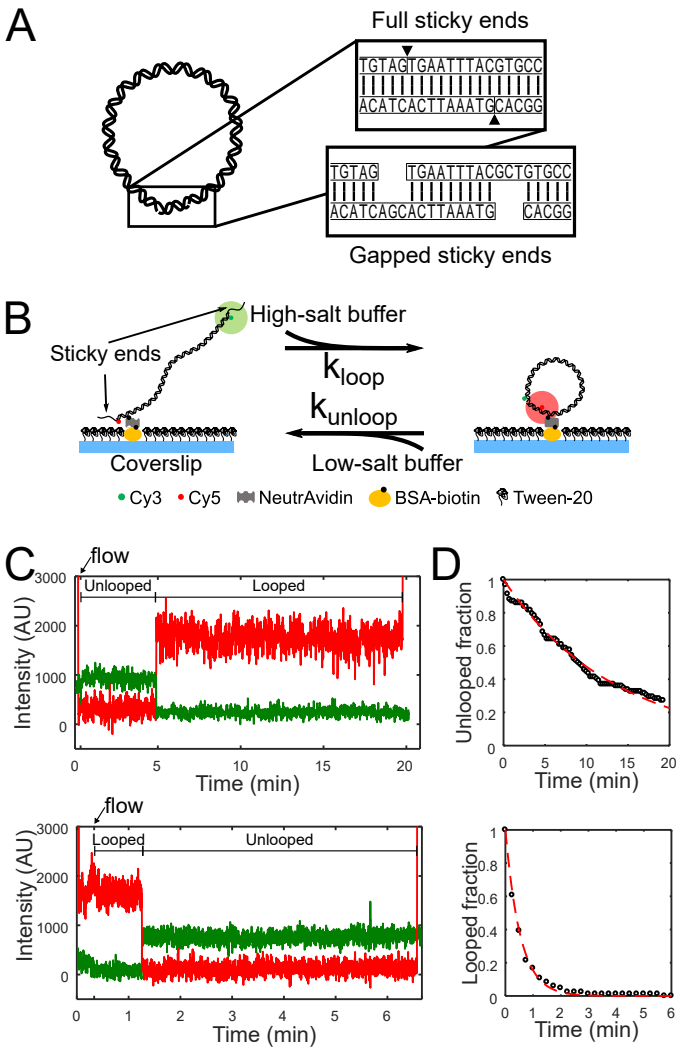


Figure 3.1: **(A)** Schematic representation of a looped DNA molecule with annealed sticky ends. Close-up views show duplexed sticky ends, which we refer to as a linker duplex, without (top) and with (bottom) gaps. **(B)** Experimental setup in the FRET-based cyclization (looping) and decyclization (unlooping) assays. Fluorescently labeled DNA molecules with sticky ends are immobilized on a passivated coverslip and continuously excited by the evanescent wave of a 532-nm laser. The cation concentration of the surrounding imaging buffer is exchanged to promote either looping or unlooping of the DNA molecules. **(C)** Examples of typical fluorescence trajectories of a single DNA molecule on the surface transitioning from the unlooped state to the looped state (top) and from the looped state to the unlooped state (bottom) upon sudden salt-exchange at time = 20 s (marked by an arrow). The green and red lines represent the donor (Cy3) and acceptor (Cy5) intensities, respectively. The molecules are briefly excited by a 640-nm laser in the beginning and the end for co-localization of Cy3 and Cy5 as well as to confirm the presence of Cy5. **(D)** Examples of decay curves of the unlooped (top) and looped (bottom) fractions of molecules. The rates are extracted by fitting the data (black) with an exponential function (red).

and biotin, see Appendix A for details) [19]. Donor-labeled and acceptor-labeled double-stranded DNA molecules were made in separate PCR reactions. The donor-labeled and acceptor-labeled molecules contained the sticky-end extension at the 5′ and 3′-end, respectively. For gapped sticky ends, a stretch of three noncomplementary bases were inserted in the extensions (Figure 3.1(A)). The donor (Cy3) and the acceptor (Cy5) were linked to the thymine bases nearest to the 5′ ends so that sticky-end annealing generated a high FRET signal ($\sim$0.8).

Strand exchange was performed between the two DNA molecules by incubating the mixture ($\sim$100 nM of Cy3-labeled DNA and $\sim$25 nM of Cy5-labeled DNA) at 95 °C for 5 minutes and gradually cooling to the room temperature. As a result of strand exchange, the majority ($\sim$70 %) of products contained all the necessary modifications as well as the 5′ protruding sticky ends.

All of the PCR primers were commercially synthesized by Eurofins MWG Operon and Integrated DNA Technology (IDT) to at least HPLC-purity grade to minimize truncation or deletion errors. We also used Mfold [63] to ensure that each sticky end does not form unintended secondary structures.

3.2.2 FRET cyclization/decyclization assay

We adopted the previous salt-exchange FRET assay [33] except for some slight modifications in the flow-cell preparation step. We started by cleaning a microscope slide with drilled holes and a coverslip by sonication in deionized water. After sonication, the slide and the coverslip were completely dried in a vacuum chamber for about 10 to 15 minutes and etched in a plasma cleaner for additional 5 minutes. A dust-free, smooth surface was obtained at this stage. Then, we silanized the slide and the coverslip in a dichlorodimethylsilane (DDS)-hexane solution as previously described [95]. After silanization, the flow-cell was assembled by joining the slide and the coverslip using double-sided tape and epoxy glue. The flow-cell was passivated and functionalized by biotinylated BSA and tween-20

before DNA molecules were injected for immobilization.

For cyclization experiments, we first incubated the molecules in an imaging buffer containing no NaCl for 10 minutes. We then started recording the time trajectories of FRET signals of the molecules and perfuse 30 μL of 1 M [NaCl] imaging buffer into the flow channel to induce looping (Figure 3.1(B)). Perfusion was controlled by a motorized syringe pump at a flow rate of 600 μL/min. The decyclization experiment was done in the same manner except that we change the salt concentration in the imaging buffer from 2 M [NaCl] to either 75 mM or 1 M [NaCl]. All imaging buffers contained the PCD-PCA oxygen scavenging system [107]. Figure 3.1(C) shows typical fluorescence intensity trajectories of Cy3 and Cy5 from these experiments. The temperature of the flow channel was maintained at 20 °C via an objective lens temperature controller at all times. Single-molecule fluorescence data were acquired on an objective-based TIR microscope with an EMCCD camera (DU-897ECS0-# BV, Andor) at a rate of 100 ms per frame.

3.2.3 Association and dissociation rates of the linker duplex

To measure the association rate (k_{on}) between the sticky ends and the lifetime (τ_{on}) of the linker duplex, we prepared four different partial DNA duplexes that are sticky on one end and blunt on the other (see Appendix A). Two of them contained full sticky ends, and the other two gapped sticky ends. Each sticky end was labeled with either Cy3 or Cy5. These partial duplexes were constructed by heating a mixture of complementary oligonucleotides to 95 °C for 5 min and gradually cooling to 4 °C. The final concentrations of the oligonucleotides were ∼10 μM. The products from this reaction were purified by native polyacrylamide gel electrophoresis (12 %, 19:1 ratio of acrylamide to bis-acrylamide in 1X TBE buffer) and extracted by “crush and soak" followed by ethanol precipitation. The concentration of the purified product was estimated from the absorbance of the fluorescent label at its maximum absorbance wavelength. To measure k_{on}, one of the partial duplexes was immobilized on the surface, the other partial duplex carrying the complementary sticky

end was injected into the flow cell at a known concentration, and the appearance of FRET events was monitored. To measure τ_{on}, linker duplexes were formed on the surface, dissociation was induced by salt exchange, and the disappearance of FRET was monitored.

3.2.4 Data analysis

We used Matlab to extract time trajectories of FRET values from the immobilized molecules. The FRET efficiency, or signal, was calculated from the background-subtracted intensities of the donor (I_D) and acceptor molecules (I_A) using $I_A/(I_A+I_D)$. The FRET time trajectories were filtered by applying a 2-point moving average and were fed to a Hidden Markov Model estimator [105] to determine the transition points between the ideal FRET levels. The first passage time to FRET transition (low to high for looping or association and high to low for unlooping or dissociation) was collected from each FRET trajectory to build the decay curve (see Figure 3.1(D)).

3.2.5 Extracting rates from the decay curves

The looping and unlooping rates were extracted from fitting an exponential function to the looping and unlooping decay curves. We observed that the unlooping decay curve reached to a zero within our typical recording time ($\sim$ 6 min), and thus used an exponential function of the form $N(t) = \exp(-k_{\text{unloop}}t)$ for fitting. However, the looping decay curve did not reach to a zero even after 20 minutes of observation, which was the longest recording time we have tried in this study. According to Ref. [54], we noted a certain fraction of DNA molecules, N_∞, did not loop even after a long time ($\sim$40 min). Therefore, we assumed that all DNA samples contain a similar fraction of inactive molecules and fitted decay curves with an equation of the form $N(t) = (1-N_\infty)\exp(-k_{\text{loop}}t)+N_\infty$, with $N_\infty = N_\infty(189\text{bp})$, which was determined from 189-bp-long DNA molecule from DNA set 2 whose decay curve quickly plateaus to $N_\infty(189\text{bp})$. The mean lifetimes, τ_{unloop} and τ_{loop}, can be found by taking the reciprocal of k_{loop} and k_{unloop}, respectively.

3.3 Results and discussion

Using the single-molecule FRET assay, we measured the cyclization and decyclization kinetics of DNA near 100 bp in length. Cyclization or decyclization was triggered by a sudden increase or decrease in NaCl concentration. The FRET signals of single molecules were continuously monitored from the beginning moment of buffer exchange, and the first transition times in the FRET signals were collected to obtain mean lifetimes or rates.

3.3.1 The looping rate changes monotonically, but the unlooping rate oscillates with DNA length

In Figure 3.2(A), we present k_{loop} of molecules in DNA set 1, decreasing in length from 116 bp to 94 bp in 2-bp steps. This range spans almost two helical periods of DNA. As shown, k_{loop} of DNA set 1 monotonically decreases as DNA becomes shorter, indicative of the increasing energy cost of looping. The difference in k_{loop} over the range of $\sim$20 bp is nearly 10-fold. The most noteworthy feature of this plot is the lack of helical-phase dependent oscillation, which shows that loop capture does not require continuity of the helical phase at the boundary. We also measured k_{loop} of DNA set 1 with gapped sticky ends that prevent stacking between opposing terminal bases [93]. As expected, DNA molecules with gapped sticky ends exhibit a similar monotonic dependence of k_{loop} on length. These molecules cyclize at a slightly slower rate due to the slower annealing rate of gapped sticky ends. These results indicate that stacking between the terminal bases of opposing sticky ends is not necessary for loop capture, and that the transition state must be torsionally relaxed. This is in contrast to the ligation-based assay which requires alignment of phosphate backbones, thus the oscillating cyclization rate with DNA length [13, 21].

Next, we present the dependence of decyclization kinetics on DNA length and sticky-end type. In Figure 3.2(B), we plot the lifetime of the looped state, τ_{loop}, which is the inverse of the decyclization rate. In contrast to cyclization kinetics, decyclization kinetics

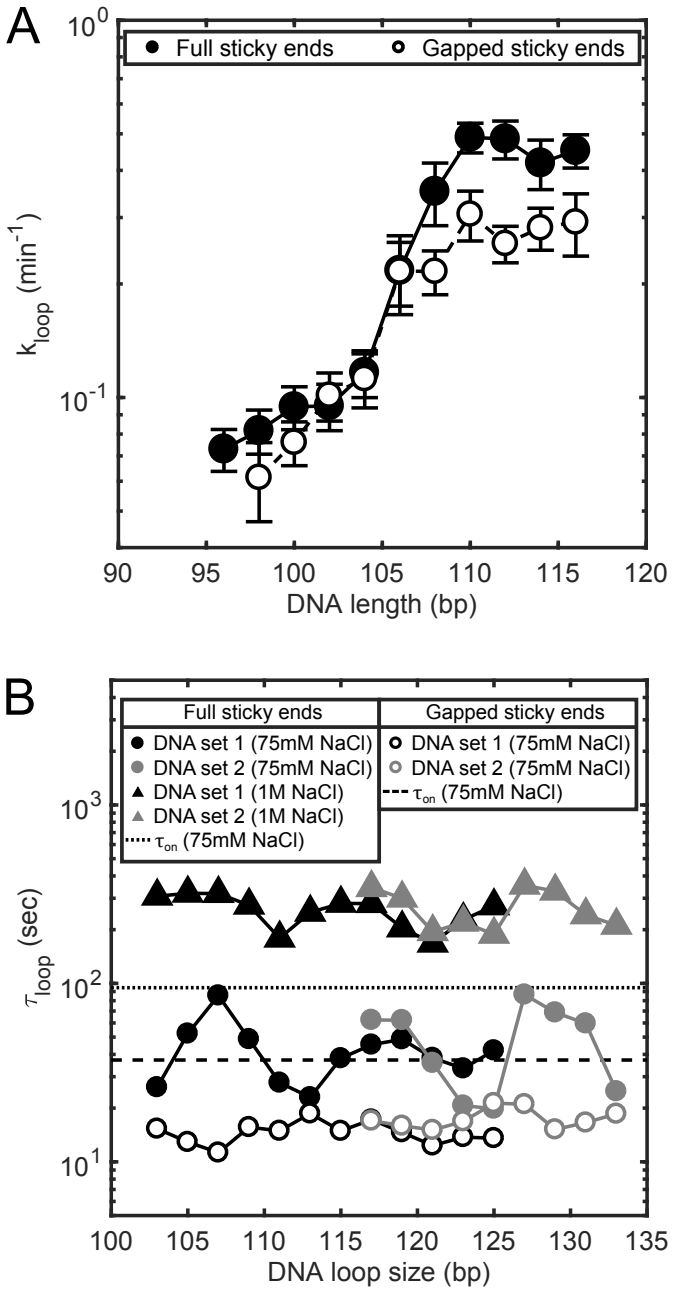


Figure 3.2: **(A)** Looping rates of DNA molecules in set 1. The closed and open circles represent data measured with the full and gapped sticky pairs, respectively. Error bars represent the standard errors of the mean. **(B)** The lifetimes of DNA molecules in DNA sets 1 and 2 in the looped state. The DNA molecules with the full sticky ends are measured in two different [NaCl] conditions: 75 mM (closed circles) and 1 M (triangles). The dotted and dashed horizontal lines represent the lifetimes of the full and gapped linker duplex, respectively. The size of error bars (not shown) is similar to the size of the data points. The DNA loop size includes the length of the annealed sticky ends (9 bp).

of DNA with full sticky ends exhibit a clear length-dependent oscillatory pattern (indicated by solid symbols). The oscillation is seen with two unrelated DNA sequences (black and gray symbols) and in two different salt conditions (circle and triangular symbols). In both salt conditions, the period of oscillation is similar to one helical period of DNA (10.5 bp). At 1 M [NaCl], local maxima is identified at $\sim$105 and $\sim$115 for DNA set 1, and $\sim$127 for DNA set 2. These values are closer to integer multiples of the helical period (105, 115.5, 126 bp) than half-integer multiples. At 75 mM [NaCl], the locations of maxima (and minima) shift towards slightly larger values, which we speculate is due to curvature-dependent unwinding of a double helix [90, 121]. The helical period (h) of a short DNA ring is predicted to be longer than the unstressed value (h_0) due to the twist-bend coupling term (B) according to [64]

$$h \approx h_0 \left(1 + \frac{1}{2}\kappa^2 \left(\frac{B}{CL}\right)^2\right), \tag{3.1}$$

where κ, C, and L are the curvature, the torsional stiffness and the contour length of DNA, respectively. We speculate that weaker electrostatic screening at lower salt increases B, thus h increases. Both h_0 and C of DNA do not depend on salt [22, 66]. In addition to the oscillation phase, salt influences τ_{loop} and its oscillation amplitude. Loops are about 10-fold longer-lived at 1 M [NaCl] (triangles) than at 75 mM [NaCl] (circles), and the oscillation amplitude is markedly larger at 75 mM [NaCl] than at 1 M [NaCl].

3.3.2 The role of base stacking in the stability of DNA loop

On the other hand, DNA loops captured with gapped sticky ends do not show length-dependent oscillation in τ_{loop} (open circles in Figure 3.2(B)). Moreover, τ_{loop} with gapped sticky ends was found to be similar in magnitude to the local minima of τ_{loop} with full sticky ends. Since the difference between full sticky ends and gapped sticky ends is the ability of base-stacking [67, 93], we reasoned that the oscillation seen with full sticky

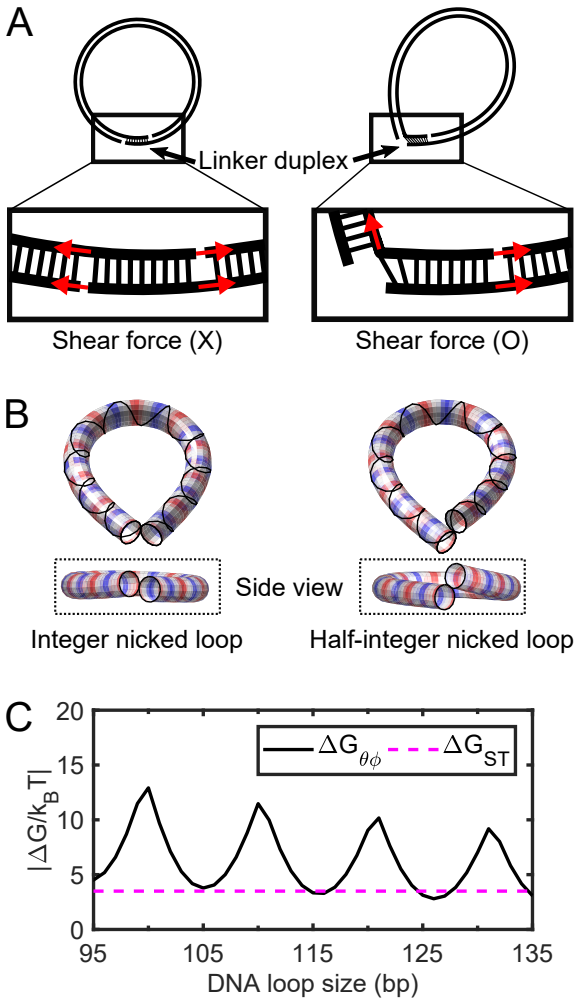


Figure 3.3: **(A)** Schematic of how nick closing (terminal base stacking) can alter the stress geometry of the linker duplex. A fully stacked linker duplex (left) does not experience a shear force and therefore is more stable. In comparison, an unstacked linear duplex (right) experiences a shear force and therefore is less stable. **(B)** Minimum-energy shapes of the coarse-grained twistable worm-like chain with a single nick (left: 105 bp and right: 100 bp). Here, we consider both ends (the first and last 10 bp) of the coarse-grained chain to be cylinders with a radius equal to 1 nm whose volumes are excluded from each other during the energy minimization procedure [70]. The strand without a nick is shown as a solid line around the tubular shapes. The alternating red and blue colors indicate one helical turn (e.g. the spacing between neighboring reds (or blue) is about one helical turn). **(C)** Comparison of free energy costs. $\Delta G_{\theta\phi}$ is the free energy cost to axially and torsionally align the ends of the helix at the tip of a small teardrop loop (solid line), and ΔG_{ST} is the average base pair stacking energy of all 16 dinucleotides taken from Ref. [111] (dashed line). The extrapolation method in the same reference is applied to extrapolate the stacking energy for 20 °C and [NaCl] = 0.1 M.

ends arises primarily from the stacking-unstacking equilibrium at the nicks in the loop; integer loops (loops with integer number of helical turns) are longer-lived than half-integer loops because of more stable base stacking. The salt-dependent changes of τ_{loop} and the oscillation amplitude are also consistent with stabilization of base stacking at the nicks [65].

We note that base stacking at the nicks can not only provide additional stability to the linker duplex, but also dramatically alter the stress geometry of the linker duplex (Figure 3.3(A)). If one or both nicks are open due to unstacking, the linker duplex would be subject to a shear stress, which accelerates melting of short DNA duplexes [73, 108, 109]. On the other hand, if both nicks are closed as a result of stacking, the linker duplex does not experience the shear stress. To test this idea, we measured the lifetime (τ_{on}) of an unstressed linker duplex produced from bimolecular association of full or gapped sticky ends. The measured lifetimes are plotted in Figure 3.2(B) as dotted are dashed lines, respectively. The amplitude of oscillation in τ_{loop} (filled circles) appears to be significantly larger than the contrast between the solid and dashed lines. This comparison reveals that stacking-dependent change in τ_{on} is not sufficient to account for the fold-change in τ_{loop} between integer and half-integer loops; therefore, the difference in the shearing geometry must be taken into account as well.

We then ask a question as to what prevents half-integer loops from stacking at the nicks. To gain insight, we compute the minimum energy conformations of integer and half-integer loops with a single open nick [70]. In this calculation, we modeled the core of DNA as a one-dimensional twisted worm-like chain and applied the end-to-end constraint to a helical strand that winds around it. As shown in Figure 3.3(B), integer loops adopt a planar teardrop shape, whereas, half-integer loops are non-planar. Therefore, nick closing which requires axial and torsional alignment at the apex of the teardrop would be energetically more challenging to half-integer loops. We can estimate the free energy cost ($\Delta G_{\theta\phi}$) for the teardrop loop to achieve axial and torsional alignment at the nick from the J factors

according to

$$\Delta G_{\theta\phi} = -k_B T \log(J_{\theta\phi}/J), \tag{3.2}$$

where $J_{\theta\phi}$ and J are the semianalytically derived J factors with and without the helical alignment, respectively [27, 104]. As predicted, $\Delta G_{\theta\phi}$ of half-integer loops is much larger than that of integer loops (solid line, Figure 3.3(C)). $\Delta G_{\theta\phi}$ of integer loop is still much larger than the thermal energy, but is comparable to the free energy (ΔG_{ST}) of base stacking (dashed line, Figure 3.3(C)), which we estimated from the literature [111]. In agreement with our thermodynamic argument, a recent coarse-grained simulation [49] also shows that half-integer loops adopt a non-planar teardrop loop configuration in which base pair stacking across nicks is disrupted.

Although indistinguishable by FRET, our kinetic analysis distinguishes two primary macrostates in the looped state. In summary, our results validate a three-state cyclization model (Figure 3.4): (1) unlooped, (2) teardrop loop (end-juxtaposed), and (3) smooth

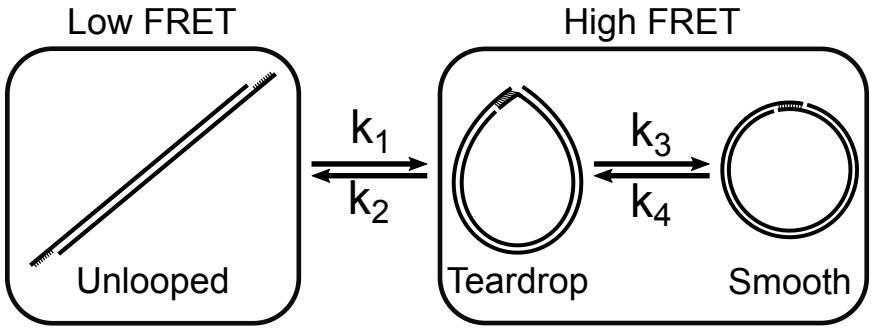


Figure 3.4: Three-state DNA cyclization model. A sticky-ended short DNA molecule undergoes a transition between the low FRET (unlooped) state and the high FRET (looped) state. The transition rates between these two FRET states (k_1 and k_2) are governed by the bending energy of DNA. Two different macrostates, teardrop and smooth, can exist within the high FRET state since the looped molecule contains nicks that can spontaneously close and open. Transitions between the teardrop and smooth states occur at the rates of (k_3) and (k_4), respectively, and are associated with local transitions of nick closing and opening. For the transition from the teardrop state to the smooth state, integer loops need axial alignment only while half-integer loops need both axial and torsional alignment. Therefore, integer loops can transition to the smooth state more readily than half-integer loops.

loop (axially and torsionally aligned, and terminally stacked). A looped state with two open nicks is also possible, but is omitted from the model because it is significantly less favourable than the other two looped states (see Appendix A.2). The oscillation-free looping rate indicates that the looped state is at first captured in a torsionally relaxed state ($k_{\text{loop}} = k_1$), which is likely a teardrop loop with an open nick(s). Thus, the transition rates (k_1, k_2) between the first two states are independent of the helical phase between the two ends. In contrast, the transition rates (k_3, k_4) between the second and third states depend on bending, twist, and stacking energies. This equilibrium explains the difference in τ_{loop} between integer and half-integer loops. Integer loops only require in-plane bending fluctuations to close the nick, while half-integer loops require energetically demanding out-of-plane deformations to do so. Therefore, half-integer loops would be stalled in the teardrop state, and decyclize at a rate of $k_{\text{unloop}} = k_2$, while integer loops would be partitioned between teardrop and smooth states, and decyclize at a slower rate of $k_{\text{unloop}} = k_2 \cdot \frac{k_4}{k_3+k_4}$. The single-exponential decay of the high FRET state implies that teardrop and smooth states equilibrate much faster than k_2. Based on our model, we propose the oscillation amplitude and phase in k_{unloop} vs. DNA length (Figure 3.2(B)) as a useful measure to probe twist-bend coupling (B) and torsional stiffness (C) of DNA in different sequence contexts or experimental conditions.

3.3.3 Revisiting the J factor of short DNA

The worm-like chain model is widely successful in describing the statistical mechanics of long DNA. However, whether it correctly describes the looping probability of DNA shorter than 100 bp is still debated. The comparison between measurement and model is most comprehensively shown on the plot of the J factor vs. DNA length, called the cyclization profile [77]. Using the single-molecule FRET assay, Vafabakhsh and Ha [15] obtained a J factor that becomes increasingly higher than the worm-like chain prediction below 100 bp. In the same study [15], the J factor was shown to oscillate in a length-dependent manner,

and DNA with half-integer helical turns had higher J factors than DNA with integer helical turns, which is quite opposite to the results of ligation-based cyclization studies [16, 17, 18].

As previously noted [55, 69], the seemingly high J factor of Vafabakhsh and Ha below 100 bp is not surprising given that their J factor (J_{VH}) was extracted from $J_{\text{VH}} = R/k_{\text{on}}$, where $R = k_{\text{loop}} + k_{\text{unloop}}$ is an apparent relaxation rate toward the equilibrium state. Hence, J_{VH} by definition is larger than $J = k_{\text{loop}}/k_{\text{on}}$ that is more closely related to the theoretical J factor. The larger discrepancy between J_{VH} and theoretical J at shorter lengths is also expected because k_{unloop} increases steeply with decreasing length [33]. Our new results from this study also offer a clear explanation to the out-of-phase oscillatory profile of J_{VH}. Although k_{loop} changes monotonically with length, k_{unloop} oscillates with peaks at half-integer helical turns. Therefore, J_{VH}, which is proportional to $k_{\text{loop}} + k_{\text{unloop}}$, would exhibit peaks at half-integer turns.

Following the correct expression of the J factor ($J = k_{\text{loop}}/k_{\text{on}}$), we extracted the J factors of DNA set 1 with full sticky ends using k_{loop} from Figure 3.2(A) and k_{on} measured from the bimolecular association of the full sticky ends. The results are shown in Figure 3.5(A) (black dots). In the same figure, we also plot the theoretical J factor of a worm-like chain [104] using a range of persistence lengths from 40 to 50 nm (dashed lines). This theoretical J factor should be taken as a lower limit as it approximates the fluctuations about the minimum-energy loop only up to the quadratic terms [119]. As shown in this plot, the J factors of DNA set 1 correspond to persistence lengths between 44 and 49 nm. This 5-nm variability is still within the accepted range of experimentally determined values [24]. We also measured J_{loop} from molecules in DNA set 2 over a wider length range (grey circles, Figure 3.5(A)). J_{loop} from these molecules shows an overall good agreement with J of 50-nm persistence length. The difference in J_{loop} between the two DNA sets is consistent, but not remarkable considering that the J factor can vary with sequence by a few orders of magnitude in the similar length range [82]. In comparison, we show that

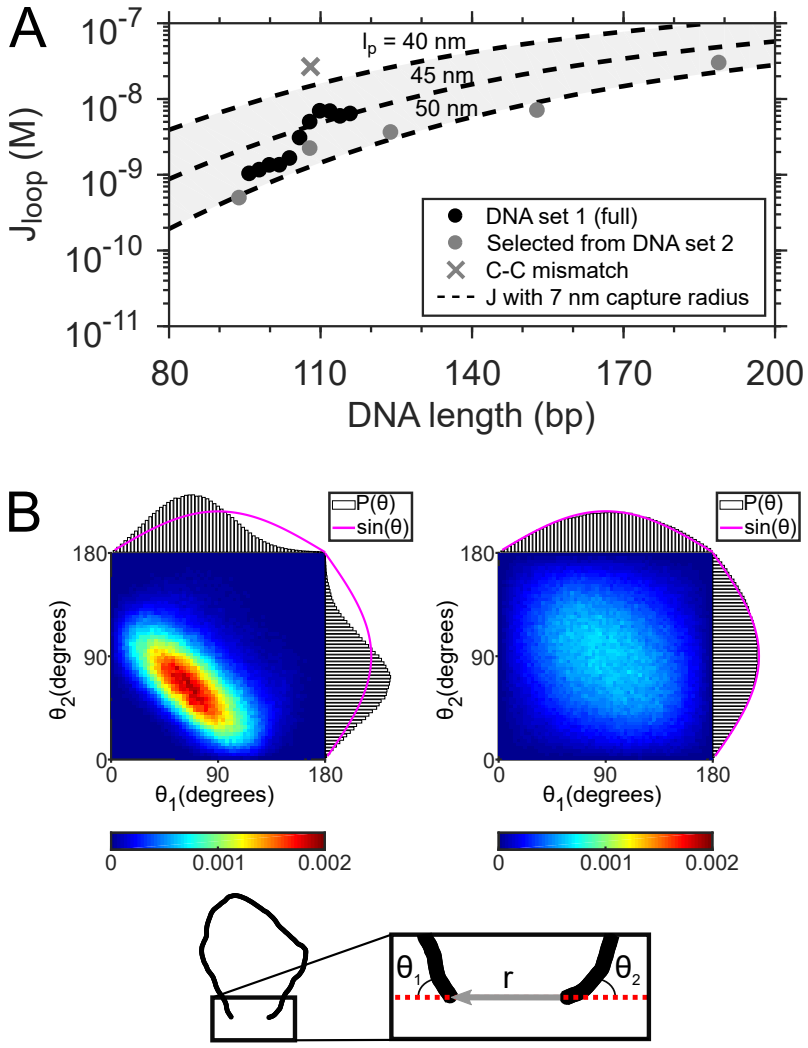


Figure 3.5: **(A)** J_{loop} as a function of DNA length. The length-dependence in the extended range outside of DNA set 1 is measured with DNA molecules from set 2 (grey). The measured J_{loop} is compared with the worm-like chain model prediction of the J factor (dashed lines) calculated based on Ref. [104]. In this calculation, we assumed the loop capture radius is equal to the contour length of 10-nt single-stranded DNA since loop capture is initiated by base pairing between single-stranded sticky ends. The shared area between the dashed lines represent the prediction made with a range of persistence lengths from 40 to 50 nm. **(B)** Joint probability distributions ($P(\theta_1, \theta_2)$) of coarse-grained DNA chains. The schematic at the bottom shows a DNA chain constrained with a short end-to-end distance, $|\mathbf{r}|$. θ_1 and θ_2 are the angles between the chain ends and the end-to-end vector. The left and right density plots represent the joint distributions of θ_1 and θ_2 for 100-bp and 500-bp loops, respectively. The projected probability distributions of θ_1 and θ_2 are individually plotted along the x- and y-axis of each density plot, respectively. The magenta line represents the unconstrained $P(\theta_1, \theta_2)$, which is equal to the sine function.

a single base pair mismatch in the center of a 108-bp DNA can increase J_{loop} by almost 10-fold (cross(x) in Figure 3.5(A)). Hence, apart from sequence-dependent irregularities, our J factor measurements are consistent with the canonical worm-like chain model in the length range tested.

3.3.4 Limitations of the J factor below 100 bp

Although understanding energetics of DNA looping at even shorter lengths ($<$100 bp) is of growing interest, we argue that the J factor is neither a theoretically relevant nor an experimentally accessible quantity in this regime. As DNA becomes shorter, end segments are more flexible than internal segments [68], and end base pairs are more prone to fraying [45, 49, 120]. Moreover, discreteness of base pairs and sequence-dependent effects cannot be sufficiently averaged out over several helical turns. Therefore, the J factor which describes the average behavior of a continuous, homogeneous polymer is no longer relevant in this length scale.

For short DNA, the experimental J_{loop} also becomes a bad proxy for the theoretical J factor. The underlying assumption in $J_{\text{loop}} = k_{\text{loop}}/k_{\text{on}}$ is that the second-order annealing rate constant between the two sticky ends in cyclization is the same as that in bimolecular association. This assumption allows the use of k_{on} measured from the bimolecular reaction to cancel out the annealing rate constant f in k_{loop} and recover the looping probability density J:

$$J_{\text{loop}} = \frac{k_{\text{loop}}}{k_{\text{on}}} = \frac{f}{k_{\text{on}}} J = J. \tag{3.3}$$

However, if the annealing rate constant depends on the relative orientation of the sticky ends, f depends strongly on DNA length. To highlight this effect, we plot the joint probability distribution of two angles ($P(\theta_1, \theta_2)$) formed between the end-to-end vector and the helical axes of the end segments (Figure 3.5(B)) using a Monte Carlo simulation of a worm-

like chain [55]. These angles thus represent how much the two sticky ends would have to deviate from the helical axes for annealing. Large angles will incur some energetic penalty because dangling bases in the sticky ends can stack [103, 106], albeit weakly. The two angles at which two separate molecules encounter would be independent and uniformly distributed, and therefore, $P(\theta_1, \theta_2)$ should be proportional to $\sin\theta_1 \sin\theta_2$ (magenta lines, Figure 3.5(B)). A similar distribution is obtained for the ends of a 500-bp DNA, much longer than the persistence length (right, Figure 3.5(B)). For the ends of short DNA (100 bp), however, the two angles are highly restrained because of the strong bending stress in the looped DNA (left, Figure 3.5(B)). Compared to the bimolecular case, small angles favorable for annealing occur more frequently while extreme angles unfavorable for annealing occur less. It is thus conceivable that at the same end-to-end distance, f would be larger than k_{on}, leading to higher J_{loop} than the theoretical J factor.

The J factor measurement of short DNA suffers from practical complications as well. As DNA becomes shorter, the fraction of molecules that loop on a laboratory time scale becomes extremely small, and detection of this trace amount in a bulk ligation assay becomes quite laborious and cumbersome [18]. In our single-molecule FRET assay, extremely slow events are inevitably masked by photobleaching of the fluorophores, which leads to an overestimation of k_{loop}. This overestimation becomes more severe as cyclization becomes slower. In our experience, slower cyclization kinetics is also fitted more poorly with a single exponential function, possibly due to an increasing inactive fraction over time. Therefore, J_{loop} of DNA shorter than 100-bp carries substantial experimental and statistical uncertainties. In our opinion, J_{loop} of short DNA should be interpreted only as a comparative measure of loopability, but not as a proxy for the theoretical J factor.

3.4 Conclusion

The single-molecule FRET assay [15, 54] can detect cyclization intermediates without the need of protein-mediated ligation and is thus thought to be a more accurate method to

measure the intrinsic looping probability of short DNA. However, exact boundary interactions and loop geometry of these intermediates are not known, which complicates the interpretation of the apparent cyclization rate. In this study, we measured cyclization and decyclization rates of short DNA as a function of DNA length. The cyclization rate changes monotonically with DNA length without helical-phase dependent oscillation. In contrast, the decyclization rate showed length-dependent oscillation, faster at half-integer helical turns, and slower at integer-helical turns. We further demonstrate that the oscillation results from stackable bases at the nicks, and the oscillation amplitude can be explained by shear-accelerated nick opening. We present a three-state cyclization model that is kinetically and thermodynamically consistent with our data and existing stacking free energy parameters, and propose the oscillation profile of the decyclization rate as a new measure to explore twist and bending mechanics of DNA. Lastly, the J factors extracted from cyclization rates of 90 to 120-bp DNA are in good agreement with persistence lengths in the range of 44 to 49 nm.

CHAPTER 4

THE EFFECT OF BASE PAIR MISMATCH ON DNA CYCLIZATION[1]

4.1 Introduction

Cellular DNA is constantly exposed to the possibility of mispairing (i.e. non-complementary base pairing)[147]. Most commonly, mismatched base pairs result from base misincorporation during gene replication[141] and heteroduplex formation between slightly different DNA sequences during homologous recombination[131]. They can also arise from exposure to DNA damaging agents that modify nucleobases[145, 149]. Due to less favorable base pairing and stacking[137], mismatched base pairs can increase local flexibility of double-stranded DNA[133, 138, 148], and consequently the capture rate of tightly bent loops[146]. For example, 1 to 3 bp-mismatch near the center of a short DNA fragment ($<$150 bp) was shown to increase the rate of DNA loop formation by one to two orders of magnitude[15, 142]. The kinetics of loop formation or capture is intuitively understood by a one-dimensional free energy curve with the end-to-end distance as a single reaction coordinate (Figure 4.1). Base pair mismatch would reduce the mechanical work required to bring two distant DNA sites to proximity, more so for a shorter end-to-end distance. Therefore, the base pair mismatch would lower the transition state relative to the unlooped state (dotted line, Figure 4.1).

Base pair mismatch is also expected to affect the breakage or release rate of small DNA loops that are captured by protein complexes[117] or by sticky ends of the DNA itself[55]. Looped DNA segments on the order of one persistence length are subject to a high level of mechanical stress; therefore, the free energy of the looped state is significantly lowered

[1]This chapter is based on a manuscript in preparation for submission as:
J. Jeong and H. D. Kim, "Base-pair mismatch can destabilize small DNA loops through cooperative kinking".

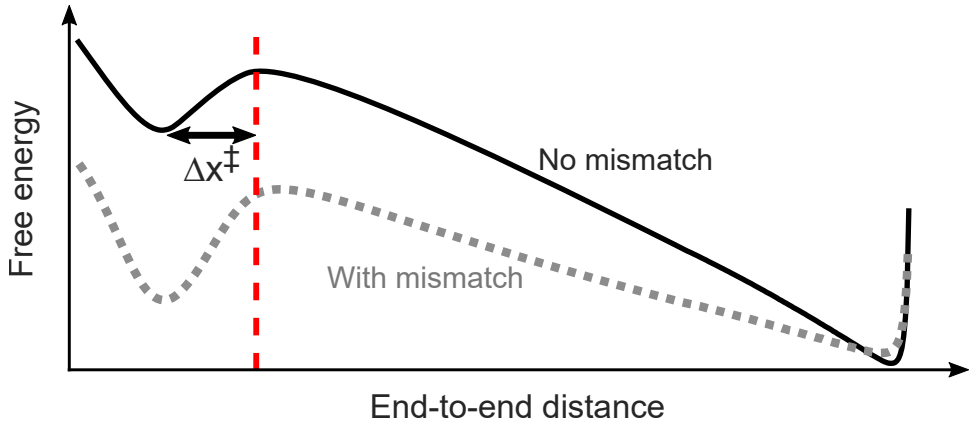


Figure 4.1: One-dimensional free energy landscape for DNA loop capture and release. The two minimum free energy states correspond to the looped and unlooped states. The transition state (vertical line) is separated from the looped state by a small distance $\Delta x^{\ddagger}$, which is equal to the capture radius. The base pair mismatch is expected to increasingly untilt the solid curve toward shorter end-to-end distances, which results in the dotted curve.

in the presence of the mismatch. According to the free energy diagram in Figure 4.1, the transition state, being at a slightly longer end-to-end distance by $\Delta x^{\ddagger}$, would be lowered to a lesser degree (Figure 4.1). Therefore, the one-dimensional model predicts that the rate of loop release would decrease in the presence of base pair mismatch.

Such prediction of mismatch-dependence seems plausible considering the success of the model in predicting the length dependence of loop capture and release rates. In the length regime where the free energy of loop formation is dominated by bending energy, increasing DNA length effectively reduces the tilt in the free energy curve because states at shorter end-to-end distances receive more stress relief, similar to the dotted line in Figure 4.1. This change predicts that loop capture and release rates measured at different DNA lengths would be anti-correlated; loops associated with higher mechanical stress are captured more slowly and released more quickly. This prediction has been confirmed for both DNA loops captured by Lac repressor[2] and DNA loops captured by sticky ends[33, 54]. While increasing DNA length evens out the bending stress over the entire DNA molecule, the base pair mismatch tends to localize sharp bending. Therefore, the effect of base pair mismatch might be quite different from that of increasing DNA length.

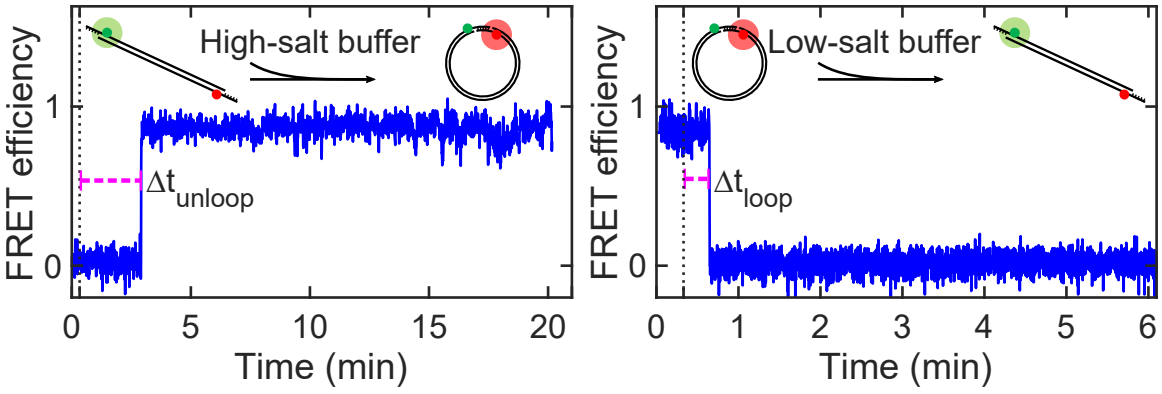


Figure 4.2: Typical FRET trajectories of a DNA molecule undergoing loop capture (left) and loop release (right). The DNA molecule labeled with Cy3 (green) and Cy5 (red) is in the low FRET state when unlooped, and in the high FRET state when looped. A sudden increase or decrease in NaCl concentration at the 20-second time point (marked by a vertical dotted line) triggers the transition.

In this chapter, we investigated how base pair mismatch affects the stability of small DNA loops. As a model system for DNA loop capture and release, we used short double-stranded DNA molecules with sticky ends. To monitor loop capture and loop release events, we used the single-molecule FRET assay as previously described in Chapters 2 and 3. Briefly, DNA molecules labeled with Cy3 and Cy5 near their sticky ends were immobilized to a NeutrAvidin-coated glass surface through a biotin linker, and loop capture or release was triggered by exchange of buffers with different NaCl concentrations (see Materials and methods).

The first transition times (Δt) in the FRET signals (Figure 4.2) of $\sim$150 individual DNA molecules were collected. The mean of Δt spent in the unlooped state before looping is defined as the loop capture time (τ_{unloop}), and the mean of Δt spent in the looped state before unlooping is defined as the loop release time or loop lifetime (τ_{loop}). All DNA molecules used in this study were shorter than 150-bp, the length regime where the free energy of loop formation is dominated by bending energy.

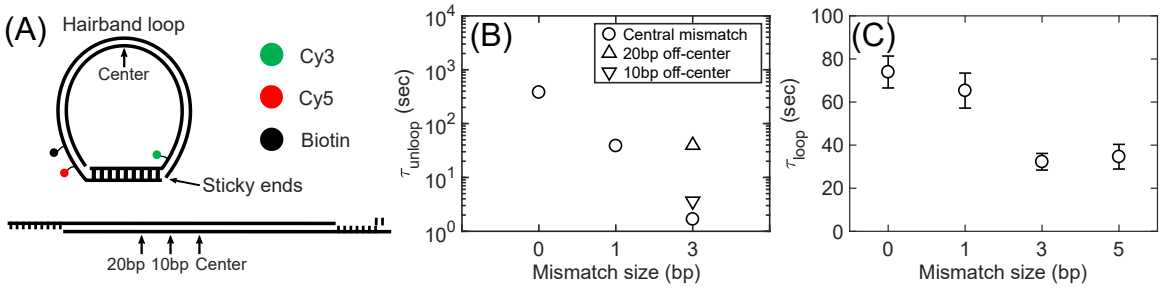


Figure 4.3: **(A)** Schematic of a hairband loop captured by sticky ends. The schematic on top shows base-paired overhangs, Cy3 (green circle), Cy5 (red circle), and the biotin linker (black circle). In this geometry, the overhangs on opposite strands form a duplex that can stack at both nicks of the loop. Different positions of base pair mismatch tested in our experiments are marked on the linear form at the bottom. Only the bases on the overhangs are shown. **(B)** Loop capture time of the hairband molecules (108 bp) as a function of the central mismatch size (circles). Data with an off-center 3-bp-mismatch are also shown as triangles. The upright and flipped triangles represent the loop capture times for base pair mismatches placed at 20 and 10 bp away from the center of the molecule, respectively. Error bars, the standard errors of the mean, are smaller than the size of the symbols. **(C)** Hairband loop lifetime (loop release time) as a function of the central mismatch size. Error bars represent the standard errors of the mean.

4.2 Results and Discussion

We first tried the loop capture geometry used in DNA cyclization, which we term as the “hairband loop" (Figure 4.3(A)). In this geometry, the complementary overhangs protrude from different strands so that the sticky ends can anneal in trans and stack upon each other. In a previous study, we showed that this end stacking, or equivalently nick closing, substantially increases the hairband loop stability[144]. Using the single-molecule FRET assay, we measured the hairband loop capture times with and without base pair mismatch in the center. As shown in Figure 4.3(B), hairband loop capture took less time in the presence of the mismatch as expected. The loop capture time further decreased with increasing mismatch size (circles, Figure 4.3(B)). The base pair mismatch in the center position led to the largest decrease in the loop capture time, and the decrease dropped as the mismatch was placed further from the center (triangles, Figure 4.3(B)). These observations confirm previous findings that mismatched base pairs reduce the energy barrier for loop formation by

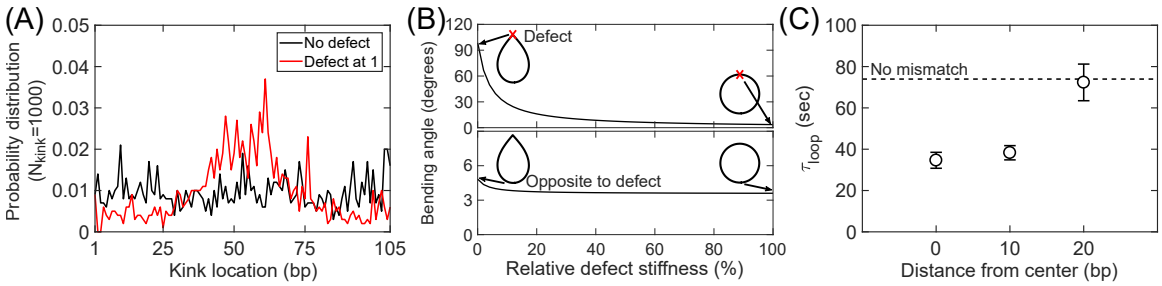


Figure 4.4: **(A)** Probability density of spontaneous kink positions along the coarse-grained minicircle (105 bp) with (red) and without a pre-existing flexible defect (black), which is placed at position 1. **(B)** Bending angle calculated from the minimum-energy conformation of a DNA minicircle (105 bp) with a defect. Top and bottom figures show bending angles at the defect and the site opposite to the defect, respectively, as a function of the defect stiffness relative to an intact base pair. The minimum-energy conformations of the two extreme cases of the defect stiffness (0 and 100%) are also shown along the curves with the defect position marked by X. **(C)** Hairband loop lifetime as a function of the mismatch position (3-bp in size). For comparison, the horizontal dotted line shows the loop lifetime without the mismatch. Error bars represent the standard errors of the mean.

increasing DNA bendability[57, 130, 133, 142], and this barrier reduction is most effective when the mismatch is in the center[39].

Next, we measured the hairband loop release times or loop lifetimes (τ_{loop}) with and without the mismatch in the center. Since a mismatch could relieve the bending stress of the hairband loop, we thought that the loop lifetime would become longer. To our surprise, we observed the exact opposite effect where the central mismatch decreased the hairband loop lifetime (Figure 4.3(C)). Increasing the size of the mismatch from 1 bp to 3 bp led to a further decrease in the lifetime. This effect seemed to plateau past the mismatch size of 3 bp (Figure 4.3(C)). This result suggests that the mismatch-containing hairband loop is more kinetically unstable than the mismatch-free loop, which seems paradoxical through the lens of the one-dimensional model presented in Figure 4.1.

We thus considered the possibility that the transition state depends on other reaction coordinates besides the end-to-end distance, such as the closing angles at the loop junction. Since base stacking at the nick(s) in the hairband loop is a key determinant of decyclization kinetics[144], we asked whether the central mismatch could destabilize the hairband loop

by allosterically inducing nick opening. To investigate such allosteric coupling, we calculated the curvature profile of a kinkable semiflexible loop[16] containing a defect with zero rigidity from a Monte Carlo simulation (see Materials and methods for details). As shown in Figure 4.4(A), a kink with a sharp bending angle appeared most frequently at the furthest end of the loop from the defect. We also calculated the minimum energy conformation of a semiflexible loop while varying the rigidity of the defect and found that the bending angles of furthest points were highly correlated (Figure 4.4(B)). This loop-mediated correlation of sharp bending angles between most distant sites is termed cooperative kinking[140], and has been observed in torsionally strained DNA minicircles by cryo-electron microscopy and molecular dynamics simulations[42, 123, 140].

We hypothesized that the enhanced flexibility of the central mismatch destabilizes the hairband loop preventing nicks(s) on the opposite side from closing. This hypothesis provides a few testable predictions. First, if the mismatch were displaced from the midpoint of the DNA, the degree of destabilization would be dampened. In agreement with this prediction, we observed a longer loop lifetime when the mismatch was placed at a quarterpoint instead of the center (Figure 4.4(C)). Second, the cooperative kinking hypothesis requires

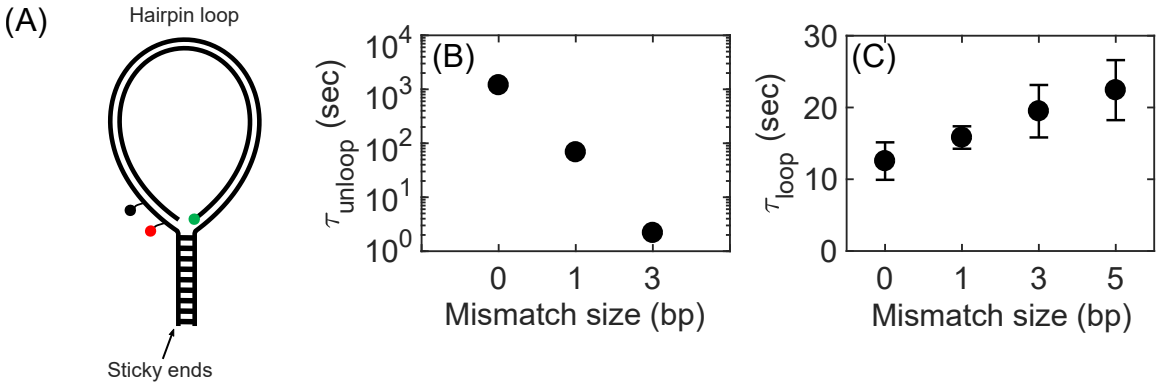


Figure 4.5: **(A)** Schematic of a hairpin loop. The schematic shows the FRET pair (green and red circles), the biotin linker (black circle), and base-paired overhangs. In this geometry, the overhangs on the same strand form a duplex like a zipper. **(B)** Loop capture time of the hairpin (105 bp) molecules as a function of the central mismatch size. Error bars are omitted due to their small sizes. **(C)** Hairpin loop lifetime as a function of the central mismatch size. Error bars represent the standard errors of the mean.

nicks that can buckle under the bending stress, and therefore the mismatch-induced destabilization would be eliminated in a loop capture geometry free of end-stacking. We thus tested a different loop geometry referred to as the “hairpin loop", where the complementary overhangs protrude from the same strand (Figure 4.5(A)). In this geometry, the sticky ends anneal in cis and cannot stack upon each other. Using these new DNA constructs with a central mismatch of various sizes, we repeated loop capture and release experiments. Similar to hairband loop capture, the hairpin capture time decreased with the size of base pair mismatch (Figure 4.5(B)). However, in sharp contrast to the hairband loop, the hairpin loop lifetime increased with mismatch size (Figure 4.5(C)). The effect of the base pair mismatch on the hairpin loop stability is therefore consistent with the prediction of the one-dimensional model. Overall, the lifetimes of hairpin loops were shorter than those of hairband loops, which is consistent with easier rupture of DNA duplex in an unzipping geometry than in a shearing geometry[129, 132, 139]. These results lend strong support to the idea that cooperative kinking governs the kinetic stability of a mismatch-containing hairband loop.

The mismatch-dependence of the hairband loop release kinetics reveals the limitations of the one-dimensional two-state model (Figure 4.1) and invites us to consider additional states and alternative reaction paths along another dimension. Here, we present two different paths ($k^{(0)}$ and $k^{(m)}$) that are likely to be the dominant ones for mismatch-free and mismatch-containing DNA (Figure 4.6(A)). Each path goes through three different states: unlooped, unstacked, and stacked. The loop capture rate is much greater in the presence of a central mismatch due to its enhanced flexibility ($k_1^{(m)} \gg k_1^{(0)}$). The reverse rate is expected to be slower with the mismatch ($k_2^{(\mathrm{m})} < k_2^{(0)}$) because of the weaker loop tension. Mismatch-free DNA undergoes small bending fluctuations uniformly throughout its contour, and therefore, follows an arc-like trajectory toward the looped state where end-stacking (nick closing) and end-unstacking (nick opening) transitions may occur. In comparison, DNA with a mismatch in the center can be sharply bent at a much lower en-

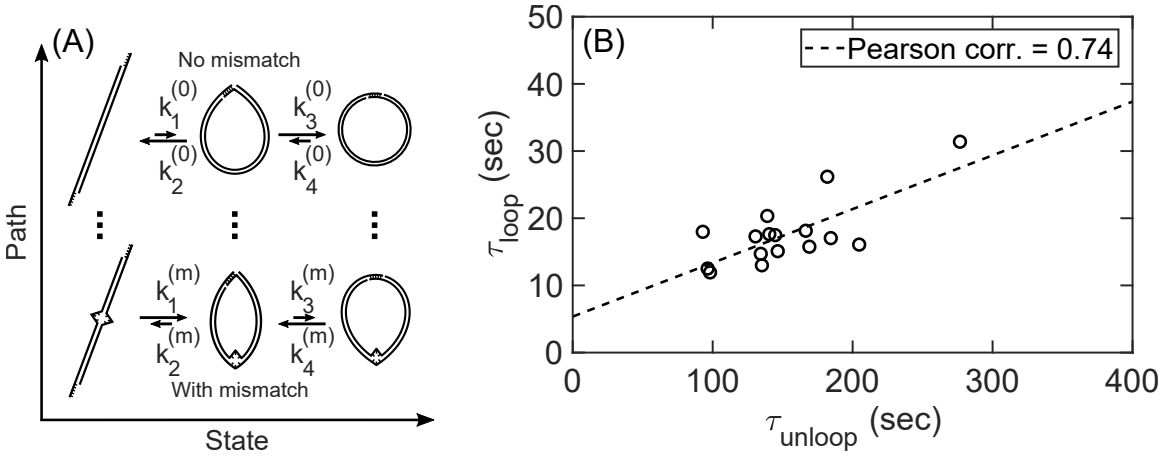


Figure 4.6: **(A)** The three-state model for hairband loop closure and release. The three states from left to right are unlooped, unstacked, and stacked states. The looped state is a mixed state between the unstacked and stacked states. Therefore, the apparent loop capture rate (k_{loop}) is equal to k_1, but the apparent loop release rate (k_{unloop}) depends on k_2, k_3, and k_4. For the hairpin loop, $k_3 = 0$, and therefore, k_{unloop} is equal to k_2. Two representative paths for central mismatch size 0 and m are highlighted with arc-like (top) and tweezers-like (bottom) motions, respectively. The vertical dotted lines imply the continuum of paths running parallel to the two extreme ones shown. **(B)** Correlation between loop capture and release times of 16 unrelated hairband DNA molecules of the same size (94bp). The loop capture and release times were measured in equilibrium (i.e. no buffer-exchange) at slightly elevated temperature of 34 °C with [NaCl] = 700mM.

ergy cost, and therefore, the most dominant path toward the looped state will resemble a tweezers-like motion. As a result of this motion, the sticky ends anneal at a sharp angle, and the hairband loop with the mismatch faces a higher energy barrier for end-stacking (nick closing) than without ($k_3^{(m)} \ll k_3^{(0)}$). The mismatch not only suppresses end-stacking, but also promotes end-unstacking (nick opening) through cooperative kinking, which implies $k_4^{(m)} \gg k_4^{(0)}$. Hence, the apparent release rate of the hairband loop (k_{unloop}) becomes faster with the mismatch than without because the looped state with the mismatch is heavily biased towards the unstacked state. In comparison, for the hairpin loop that cannot proceed to the stacked state, the three-state model is reduced to the two-state model, and the loop release rate is slower with the mismatch ($k_2^{(\text{m})} < k_2^{(0)}$).

The two paths drawn in Figure 4.6(A) represent the two most extreme paths in terms of kinetics, the top path for the slowest hairband loop capture and release, and the bottom

for the fastest. In reality, there exists a continuum of paths going through the three states with intermediate rates, and the flexibility profile of DNA determines the relative weights at which individual paths are taken. Therefore, any changes to the flexibility profile of DNA would lead to correlated changes in the hairband loop capture and release rates. To test this idea, we measured hairband loop capture and release times of 16 unrelated sequences, all of the same length. Although limited in sample size, we observed a significant degree of correlation between the two times (Pearson correlation = 0.74, Figure 4.6(B)). This result suggests that cooperative kinking is a general mechanism that governs the kinetics of hairband loop capture and release.

4.3 Conclusion

In conclusion, we demonstrate that base pair mismatch can constrain the geometry and interactions for DNA loop capture through cooperative kinking, and the close coupling between hairband loop geometry and end-stacking can give rise to correlated changes between loop capture and release times ("easy come, easy go"). We propose a three-state model that correctly describes the effect of mismatched base pairs on the apparent kinetics of loop capture and release. We expect the effect of mismatched base pairs on protein-mediated DNA loops to be more complex because of the diversity in loop capture geometry[124]. Beyond passively captured DNA loops, it would be interesting to investigate whether base pair mismatches can also influence the kinetics of DNA loop extrusion[125, 126] through cooperative kinking.

4.4 Materials and methods

4.4.1 Preparation of DNA molecules

A 105-bp-long DNA molecule was extracted from yeast genomic DNA by polymerase chain reaction (PCR) to serve as a control DNA template without any structural defect.

To probe the effect of a permanent defect, we planned to introduce a DNA mismatch to the control molecule by mixing it with its mutant followed by a strand-exchange reaction. To do so, we additionally prepared a set of mutated DNA molecules that differ from the control only in a certain location in which we put a mutation of size equal to 1bp, 3bp, or 5bp. To make such molecules, first, the mutated templates of the control DNA were synthesized from Eurofins Genomics (EXTREMer oligos) and duplexed via PCR. Each of the duplexed products was then incorporated into a pJET1.2\blunt vector (ThermoFisher) and cloned into DH5α *Escherichia coli* cells. Finally, the cloned fragments of DNA were extracted via colony PCR from the cells and were sequenced to ensure the correct mutation was made at the desired location.

To modify these molecules to carry a FRET pair (i.e. Cy3 and Cy5), biotin, and single-stranded sticky ends, we followed our standard preparation protocol [19], which involves a series of PCR and strand exchange reactions that can be found in elsewhere. For introducing a DNA mismatch in the final construct, we mixed the Cy3-labled control molecule with one of the Cy5-labeled mutated molecules with a ratio of 4:1 in the strand-exchange reaction.

The final DNA construct generated by this protocol carries a $5'$ protruding sticky end on each end and makes a hairband loop upon end-annealing as shown in Fig. 4.3(A) of the main text. We also made hairpin loops by having sticky ends on the same DNA strand (Fig. 4.5(A) of the main text).

A complete list of all DNA sequences can be found in Tables 4.1 and 4.2. Both top ($5'$ to $3'$) and bottom ($3'$ to $5'$) sequences are shown. The underlined sequences represent sticky ends. A Cy5 fluorophore is internally attached at the thymine base colored in red. A Cy3 fluorophore is either at the green thymine base or the $5'$ end of the bottom strand. A biotin molecule is linked to the thymine base shown as [T]. Hairpin molecules includes a 2-nt gap (indicated by sequences in parentheses) near each end of the top strand before sticky ends.

Hairband DNA molecules	
No mismatch	5′−TGAATTTACGTGCCAGCAACAGA[T]AGCCGCGATCGC CATGGCAACGAGGTCGCACACGCCCCACACCCAGACCT CCCTGCGAGCGGGCATGGGTACAATCATTCGAGCTCGTT GTAG-3′ 3′−CACGGTCGTTGTCTATCGGCGCTAGCGGTACCGT TGCTCCAGCGTGTGCGGGGTGTGGGTCTGGAGGGA CGCTCGCCCGTACCCATGTTAGTAAGCTCGAGCAAC ATCACTTAAATG-5′
1bp-mismatch (central)	5′−TGAATTTACGTGCCAGCAACAGA[T]AGCCGCGATCGC CATGGCAACGAGGTCGCACACGCCCCAGACCCAGACCT CCCTGCGAGCGGGCATGGGTACAATCATTCGAGCTCGTT GTAG-3′ 3′−CACGGTCGTTGTCTATCGGCGCTAGCGGTACCGT TGCTCCAGCGTGTGCGGGGTCTGGGTCTGGAGGGA CGCTCGCCCGTACCCATGTTAGTAAGCTCGAGCAAC ATCACTTAAATG-5′
3bp-mismatch (central)	5′−TGAATTTACGTGCCAGCAACAGA[T]AGCCGCGATCGC CATGGCAACGAGGTCGCACACGCCCCGGGCCCAGACCT CCCTGCGAGCGGGCATGGGTACAATCATTCGAGCTCGTT GTAG-3′ 3′−CACGGTCGTTGTCTATCGGCGCTAGCGGTACCGT TGCTCCAGCGTGTGCGGGGCCCGGGTCTGGAGGGA CGCTCGCCCGTACCCATGTTAGTAAGCTCGAGCAAC ATCACTTAAATG-5′

Continued on next page

Table 4.1 – continued from previous page

5bp-mismatch (central)	5′−TGAATTTACGTGCCAGCAACAGA[T]AGCCGCGATCGC CATGGCAACGAGGTCGCACACGCCCGCGCGCCAGACCT CCCTGCGAGCGGGCATGGGTACAATCATTCGAGCTCGTT GTAG-3′ 3′−CACGGTCGTTGTCTATCGGCGCTAGCGGTACCGT TGCTCCAGCGTGTGCGGGCGCGCGGTCTGGAGGGA CGCTCGCCCGTACCCATGTTAGTAAGCTCGAGCAAC ATCACTTAAATG-5′
3bp-mismatch (10 bp off-center)	5′−TGAATTTACGTGCCAGCAACAGA[T]AGCCGCGATCGC CATGGCAACGAGGTCGTGGACGCCCCACACCCAGACCT CCCTGCGAGCGGGCATGGGTACAATCATTCGAGCTCGTT GTAG-3′ 3′−CACGGTCGTTGTCTATCGGCGCTAGCGGTACCGT TGCTCCAGCACCTGCGGGGTGTGGGTCTGGAGGGA CGCTCGCCCGTACCCATGTTAGTAAGCTCGAGCAAC ATCACTTAAATG-5′
3bp-mismatch (20 bp off-center)	5′−TGAATTTACGTGCCAGCAACAGA[T]AGCCGCGATCGC CATGGCGGTGAGGTCGCACACGCCCCACACCCAGACCTC CCTGCGAGCGGGCATGGGTACAATCATTCGAGCTCGTTG TAG-3′ 3′−CACGGTCGTTGTCTATCGGCGCTAGCGGTACCGC CACTCCAGCGTGTGCGGGGTGTGGGTCTGGAGGGA CGCTCGCCCGTACCCATGTTAGTAAGCTCGAGCAAC ATCACTTAAATG-5′

Table 4.1: Hairband DNA molecules.

Hairpin DNA molecules	
No mismatch	5′−TGAATTTACG(CT)GTGCCAGCAACAGA[T]AGCCACAT CGCCATGGCAACGAGGTCGCACACGCCCCACACCCAGA CCTCCCTGCGAGCGGGCATGGGTTGCATGTCAGCTATGG ATCCATTCGTAAATTCA-3′ 3′−CACGGTCGTTGTCTATCGGTGTAGCGGTACCGTTGCT CCAGCGTGTGCGGGGTGTGGGTCTGGAGGGACGCTCGC CCGTACCCAACGTACAGT(CG)ATACCTAGGT-5′[Cy3]
1bp-mismatch (central)	5′−TGAATTTACG(CT)GTGCCAGCAACAGA[T]AGCCACAT CGCCATGGCAACGAGGTCGCACACGCCCCAGACCCAGA CCTCCCTGCGAGCGGGCATGGGTTGCATGTCAGCTATGG ATCCATTCGTAAATTCA-3′ 3′−CACGGTCGTTGTCTATCGGTGTAGCGGTACCGTTGCT CCAGCGTGTGCGGGGTCTGGGTCTGGAGGGACGCTCGCC CGTACCCAACGTACAGT(CG)ATACCTAGGT-5′[Cy3]
3bp-mismatch (central)	5′−TGAATTTACG(CT)GTGCCAGCAACAGA[T]AGCCACAT CGCCATGGCAACGAGGTCGCACACGCCCCGGGCCCAGA CCTCCCTGCGAGCGGGCATGGGTTGCATGTCAGCTATGG ATCCATTCGTAAATTCA-3′ 3′−CACGGTCGTTGTCTATCGGTGTAGCGGTACCGTTGCT CCAGCGTGTGCGGGGCCCGGGTCTGGAGGGACGCTCGC CCGTACCCAACGTACAGT(CG)ATACCTAGGT-5′[Cy3]

Continued on next page

Table 4.2 – continued from previous page

5bp-mismatch (central)	5′−TGAATTTACG(CT)GTGCCAGCAACAGA[T]AGCCACAT CGCCATGGCAACGAGGTCGCACACGCCCGCGCGCCAGA CCTCCCTGCGAGCGGGCATGGGTTGCATGTCAGCTATGG ATCCATTCGTAAATTCA-3′ 3′−CACGGTCGTTGTCTATCGGTGTAGCGGTACCGTTGCT CCAGCGTGTGCGGGCGCGCGGTCTGGAGGGACGCTCGC CCGTACCCAACGTACAGT(CG)ATACCTAGGT-5′[Cy3]

Table 4.2: Hairpin DNA molecules.

4.4.2 single-molecule FRET looing and unlooping assay

We followed our previous single-molecule FRET assay that employs the sudden salt-exchange protocol [33, 144]. For cyclization, DNA molecules were deposited on a passivated surface of a flow-cell and were incubated at a low salt (10 mM [NaCl]) imaging buffer containing the PCD-PCA oxygen scavenging system [107] for 10 minutes. We then injected a high salt (1 M [NaCl]) imaging buffer into the flow-cell to promote sticky ends to capture the loop configuration. Decyclization measurements were done similarly, except that the NaCl concentration was changed from 2 M to 75 mM. The immobilized molecules were excited by a 532-nm laser continuously through an objective-type TIR microscope from the beginning of the buffer exchange. The time trajectories of FRET signals (Fig. 4.2 of the main text) from the molecules were recorded by an EMCCD camera (DU-897ECS0-# BV, Andor) at a rate of 100 ms per frame for the mismatch-free molecules and 50 ms per frame for the molecules with a mismatch.

4.4.3 Minicircle simulations

The Monte Carlo simulation of a minicircle was implemented as previously described [33, 128]. A set of 105 connected nodes was used to create a coarse-grained representation of a DNA minicircle of 105 bp. The bending energy at each node was described by the kinkable worm-like chain model [128] with the parameters of b = 0.3 and h = 12 following the same notation used in Ref. [17]. We performed the simulation with and without a flexible defect of zero bending energy placed at a fixed location. For the case of no flexible spot, we first initialized the simulation without allowing the kink formation. Once the kink-free simulation was equilibrated, we allowed spontaneous kinks to appear. To construct the probability density of kink positions, we ran the simulation and stop at the first appearance of a kink. We then recorded the position of this kink and equilibrated back to the kink-free state. This procedure was repeated until we collected a distribution of 1000 kink positions. The same procedure was repeated in the presence of the hyperflexible spot to predict the effect of a flexible spot on the probability distribution of kink.

CHAPTER 5

FUTURE PROSPECTS

DNA cyclization is not only useful for studying intrinsic DNA bendability, but also for other biophysical mechanics of DNA. As a continuation of this work, I will present a few experimental studies using cyclized DNA to address questions related to (1) mechanics of short DNA duplex formation and dissociation under tension and (2) sequence-dependent mechanical properties of mismatched DNA.

5.1 Rupture of short DNA duplex at low force

5.1.1 Introduction

The formation and dissociation of a short DNA duplex (less than 30 bp) are widely exploited in many areas, such as molecular biology and DNA nanotechnology. Applications relying on short duplex formation and dissociation include PCR [152], fluorescence in situ hybridization [153], molecular force sensors [155], DNA computing [154], and DNA-based nanofabrication [156]. In fact, I have demonstrated one such application in this thesis. As discussed in the previous chapters, DNA cyclization experiments also rely heavily on the formation and dissociation of a short duplex (i.e., sticky ends) holding the two ends of DNA. In addition, DNA inside a cell frequently undergoes duplex separation and reannealing during important processes such as transcription [157] and replication [158]. Understanding the mechanism and determinants underlying duplex formation and dissociation could therefore provide insights regarding the thermodynamic and mechanical stability of DNA, as well as DNA manipulation by motor proteins inside the cell.

Most insights concerning the thermodynamic and mechanical properties of short DNA come from single-molecule force spectroscopy experiments using atomic force micro-

scopes, magnetic tweezers, or optical tweezers. These experiments typically measure force-dependent duplex separation transitions of DNA, which could be tested against polymer theories such as the worm-like chain (WLC) model [102].

One very insightful study is the recent experiment by Whitley et al. [73] in which the association and dissociation kinetics of a short DNA duplex under a stretching force were investigated using a novel approach involving optical tweezers combined with fluorescence illumination [159]. In their study, Whitley et al. tethered each end of a single-stranded DNA (ssDNA) molecule to a double-stranded DNA handle attached to a glass bead trapped in optical tweezers, such that the tension along the ssDNA molecule could be manipulated. They then flowed in complementary oligonucleotide fluorescent probes that could anneal onto the tethered ssDNA, allowing monitoring of the association and dissociation kinetics of the probe. Whitley and colleagues found that although the association kinetics were not dependent on the stretching force across the target, the dissociation kinetics were highly force-dependent. Overall, increasing the force increased the dissociation rate within the range of forces tested. Between $\sim$5 pN and 20 pN, this force dependence of the dissociation rate generally followed an exponential function, which is in accordance with Bell's theory[53]. At a force regime below $\sim$5 pN, however, the exponential dependence started to fade, eventually exhibiting what they called the "roll-over" behavior, in which the dependence on force reverses; that is, stretching at lower forces stabilizes the duplex instead of separating it. Because their optical tweezer setup was not suitable for the low force application, whether or not this behavior extended to the zero-force condition was not confirmed experimentally. Nonetheless, Whitley et al. were able to provide a plausible explanation for this non-exponential behavior by modifying Bell's model, incorporating force dependence into the elasticity of a DNA duplex in the transition state, in terms of the stretching coordinate. To fully validate this observation and their model, however, experimental evidence under extremely low force conditions appears to be necessary.

Motivated by that experiment, we devised an experiment based on using cyclized DNA

as a force applicator to generate stretching forces across a short DNA duplex and measure the force dependence of duplex formation and separation kinetics. The idea here is to generate a series of cyclized DNA of varying sizes with ends that are covalently linked by a ssDNA labeled with a fluorophore – for instance, a FRET donor. This idea is conceptually similar to DNA-based force probes demonstrated by Shroff et al. [80], Zocchi and his colleagues [160], and Mustafa et al. [85], but here we employ the ssDNA linker, as in Whitley et al.'s experiment [73], as a target sequence that will be bound by a complementary probe molecule labeled with a matching FRET acceptor to the substrate within the cyclized construct. The range of varying loop sizes would generate a range of stretching forces on the picoNewton scale, down to the entropic force limit of a long DNA polymer (for instance, see Figure 5.4(B)) [150]. The stretching force by cyclized DNA can be fairly accurately estimated based on the WLC prediction [80], and can further be calibrated by complementary force spectroscopic techniques, such as magnetic tweezers, if necessary. If the purpose, however, is to merely capture the force-dependence of the rate in the low force regime from no force to a few pN, the WLC estimation should be sufficient without such calibration. Moreover, using cyclized DNA as a molecular force generator offers several advantages over other force spectroscopy methods. First, the cyclized DNA force generator can be surface tethered and illuminated by a simple TIRF microscope, which means less optimization and troubleshooting in terms of instrumentation. The use of TIRF microscopy also implies the same setup can be used for measuring zero-force association and dissociation kinetics on the surface, which may not be done easily by other means, such as optical tweezers. Owing to surface tethering, this method also provides higher throughput than the other means.

5.1.2 Materials and methods

As briefly mentioned above, this method involves two different DNA molecules: a cyclized DNA linked by a ssDNA substracte (i.e., a partial minicircle), which is the target substrate,

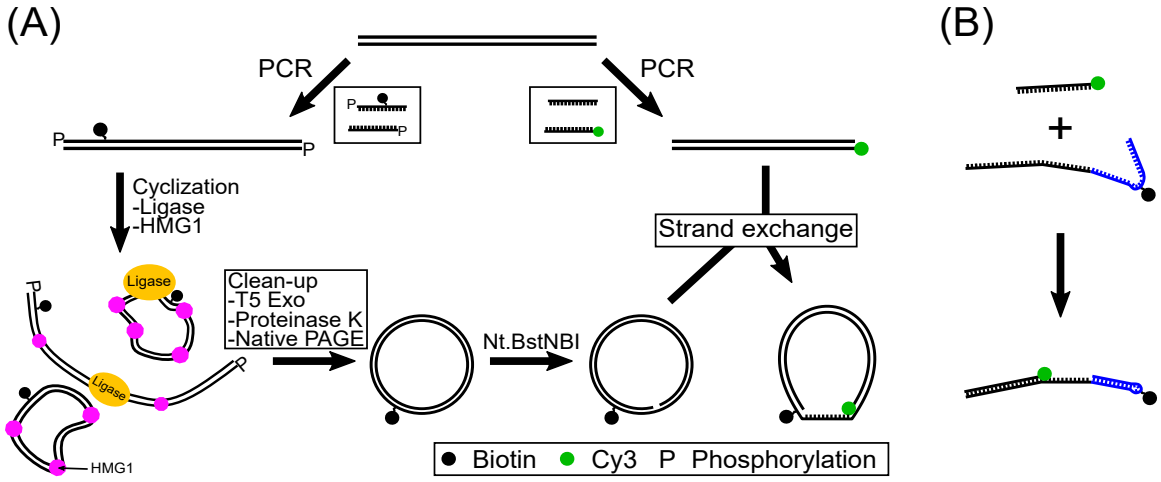


Figure 5.1: Preparation of cyclized DNA force generators. **(A)** DNA molecules are first phosphorylated by PCR with phosphorylated primers. The phosphorylated DNA molecules are mixed with ligase and DNA bending protein HMG1. As a result, ∼30 to 40% of the linear DNA molecules are cyclized. The ligation reaction is stopped by heat inactivation. The ligation products are subsequently treated with T5 exonuclease and Proteinase K. The reaction products are analyzed by native PAGE and circular monomers are extracted from the gel. The circular molecules are further treated with nicking enzyme Nt.BstNBI. Meanwhile, Cy3-label linear DNA molecules are prepared by a separate PCR reaction. The PCR products and the nicked circular molecules are mixed at a ratio of 4:1 (linear:circular) for a strand exchange reaction. After the strand exchange reaction, partially duplexed, permanently cyclized DNA molecules are made. Only the ones with modifications (i.e., biotin and Cy3) will be used for the FRET experiment. **(B)** Control DNA constructs with no DNA loop are made from directly annealing two ssDNA sequences, one with Cy3 and the other with biotin. The biotinylated DNA makes a partial hairpin indicated by blue. The annealing reaction result in partially duplexed DNA molecules with ssDNA in the middle, whose sequence (15 nt in length) is identical to the single-stranded region of the circular construct.

and a short ssDNA probe complementary to the sequence of the ssDNA target. We assume that a stretching force would be applied across the target due to the strong bending of the duplex segment. This setup allows us to measure the association and dissociation reactions between the ssDNA probe and the target, as a function of the loop size using FRET.

A series of such cyclized DNA of varying duplex sizes was constructed first by constructing DNA minicircles via ligation-based cyclization of short linear DNA fragments aided by DNA bending proteins. We then replaced one strand of the circular product with a slightly shorter complementary linear DNA strand via a strand-exchange reaction, expos-

ing the target sequence as single-stranded in the final DNA construct. This procedure is visually summarized in Figure 5.1(A).

DNA preparation

We began by PCR-amplifying a set of linear DNA molecules 84 bp, 126 bp, and 158 bp in length with phosphorylated primers (Table 5.1), using yeast genomic DNA as a PCR template. Several modifications were made simultaneously in this PCR reaction. First, we introduced some modifications to the sequence in the forward PCR primer such that one end of the PCR product carried both a nicking restriction site and a 15-nt extra tail, which would later serve as a ssDNA target in the final construct. The forward primer also carried a biotin attached at one of the thymine bases of the sequence for surface immobilization. After the PCR reaction, we tested the reaction by agarose gel electrophoresis and gel-purified the PCR products.

DNA sequences (5′ to 3′)	
158 bp	GACTCCCCACTCGTCGTACTGTTTTCCCCAGGCCAGTGCTTTAG CGTTAACTTCCGGAGCCACACCGGTGCAAACCTCAGCAAGCAG GGTGTGGAAGTAGGACATTTTCATGTCAGGCCACTTCTTTCCGG AGCGGGGTACCTCTGACTTGAGCGTCG
126 bp	GACTCCCCACTCGTCGTACGTGCTTTAGCGTTAACTTCCGGAGC CACACCGGTGCAAACCTCAGCAAGCAGGGTGTGGAAGTAGGAC ATTTTCATGTCAGGCCACTTACCTCTGACTTGAGCGTCG
84 bp	GACTCCCCACTCGTCGTACAGCCACACCGGTGCAAACCTCAGC AAGCAGGGTGTGGAAGTAGGATACCTCTGACTTGAGCGTCG
PCR primer pairs for cyclized DNA (5′ to 3′)	
Continued on next page	

Table 5.1 – continued from previous page

Forward	[Phos] TTTTGAATTTACTTTGACTCCCCAC[BiotindT]CGTCGTAC
Backward	[Phos] CTCAGTGCTGAGGTACCAGG
PCR primer pairs for linear DNA (5′ to 3′)	
Forward	GACTCCCCACTCGTCGTAC
Backward	[Cy3]CTCAGTGCTGAGGTACCAGG
Hairpin-forming strand for control DNA (5′ to 3′)	
CCTGGTACCTCAGCACTGAGTTTTGAATTTACTTTGACTCCCACC GTCGT[BiotindT]TTCGACGGTGGGAGTC	

Table 5.1: The forward primer for cyclized DNA contains the 15-nt ssDNA sequence (red) and the target site (underlined). The control DNA (see Figure 5.1B) is made by annealing between the backward primer for linear DNA and the hairpin-forming strand, which also contains the identical 15-nt ssDNA (red) as in the cyclized DNA construct.

HMG1-assisted DNA cyclization

To efficiently cyclize the phosphorylated PCR products of different sizes, we incubated 250 µg of each PCR copy in a T4 ligase reaction buffer with the DNA bending protein HMG1 (Sigma-Aldrich) at 10°C for 15 minutes to promote formation of tightly folded molecules, which would increase the probability of intramolecular end-ligation [100, 101]. The concentration of HMG1 was determined by keeping a 50:1 HMG1 to DNA molar ratio. After the incubation, we added T4 ligase into the reaction and further incubated at 15°C overnight. The ligation reaction was then stopped via heat inactivation and treated with T5 Exonuclease (10 U) to digest all the linear byproducts. At the end of the reaction, we used Proteinase K (0.5 µg) to remove the protein leftovers in the reaction, and the cyclized DNA product was concentrated by ethanol precipitation. The cyclization reaction of all three sizes (84 bp, 126 bp, and 158 bp) was confirmed by native polyacrylamide gel electrophoresis (PAGE) on a 6% gel with a 29:1 acrylamide to bis-acrylamide ratio in 1X

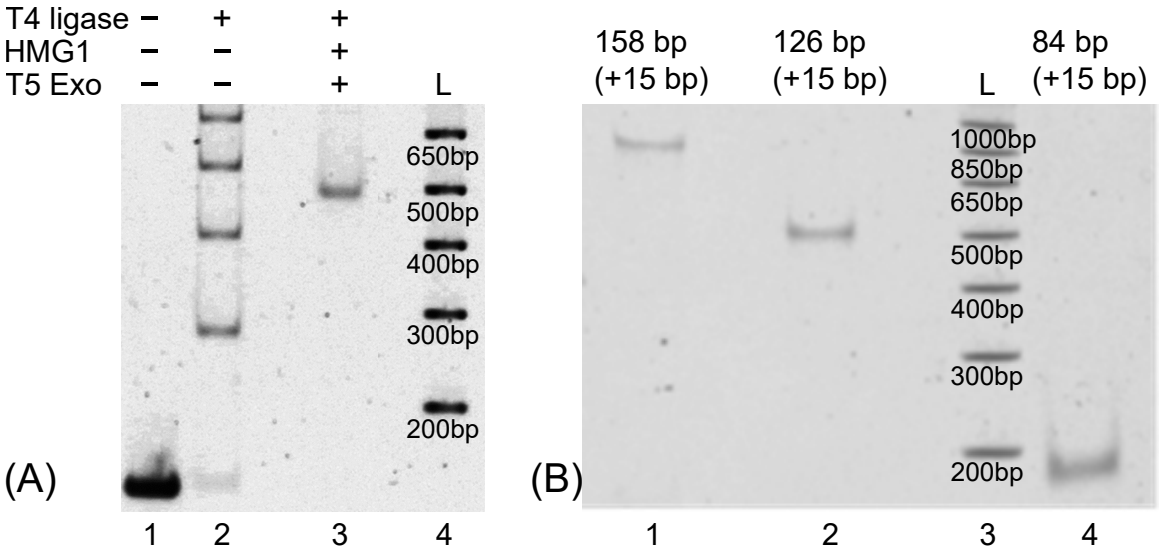


Figure 5.2: **(A)** Image of stained native PAGE comparing ligation (Lane 2) and HMG1-assisted cyclization (Lane 3) of the phosphorylated 126-bp DNA sample. Without HMG1, the ligation reaction only produced linear multimers of the phosphorylated DNA sample. T5 Exonuclease was used to remove all linear and nicked DNA byproducts of the cyclization reaction. The actual length of the DNA sample includes the 15-bp target sequence. Lane 1 shows the migration of the DNA sample without any enzyme. Lane 4 represents 1 kb-Plus ladder. **(B)** Image of stained native PAGE showing cyclized products of DNA in three different lengths. Lanes 1, 2, and 4 show the migration of 158 bp, 126 bp, 84 bp DNA samples, respectively. The actual length of the DNA samples includes the 15-bp target sequence. Lane 3 represents 1 kb-Plus ladder.

TBE (see Figure 5.2). The circular monomer product was extracted from the PAGE gel using the "crush-and-soak" method, and the extracted DNA was precipitated by ethanol for further modification.

Construction of the cyclized DNA force applicator

To incorporate the ssDNA target sequence within the double-stranded DNA minicircle, we separately prepared linear DNA with sequences complementary to the sequences of the minicircle products, minus the 15-nt target sequence, via PCR followed by gel-purification. A Cy3 molecule was attached at the 5′ end of the forward PCR primer (Table 5.1). To introduce the single-stranded target sequence in the minicircle, we performed a similar strand exchange reaction by mixing and heating the duplexed minicircles and the Cy3 labeled

linear molecules with a circular DNA to linear DNA ratio of 1:4, as in the sticky-ended DNA preparation described in the previous chapters. To allow strand exchange, however, the minicircles had to be nicked. For nicking, we treated the minicircle product with Nt.BstNBI, an enzyme that leaves the strand carrying the biotin as an unnicked circle while effectively nicking the complementary strand. Thus, the nicked circular product became completely meltable during the strand exchange reaction.

Construction of the force-free control DNA substrate

As a control experiment with no stretching force, we prepared a partially-duplexed linear DNA construct with a single-stranded region near the center by annealing two synthetic ssDNA (see Figure 5.1(B) and Table 5.1). Each of the synthetic ssDNA was purchased from Eurofins. The annealing reaction mixture contained 10 μM of each strand. One of the ssDNA strands was the Cy3-labeled backward primer. The other was a hairpin-forming strand and contained biotin. The sequence of the single-stranded region within the final construct was designed to be identical to that inside the circular construct.

FRET measurement

The FRET measurement was conducted similarly to the unlooping assay described previously for both circular and linear constructs. For this experiment, the DNA constructs (either circular or linear) were immobilized on the surface of a passivated flow-cell via the Neutravidin-biotin interaction (Fig 5.3(A)). By design, only correctly labeled (i.e., with Cy3 and biotin) final constructs were allowed to be immobilized and to fluoresce. We observed association and dissociation events of the Cy5-labeled ssDNA probe (directly purchased from Eurofins) on these immobilized substrates in equilibrium, provided that a PCD-PCA based imaging buffer containing 50 mM NaCl as well as 20 nM of the ssDNA probe was present in the flow-cell. A high Cy5 (or high FRET) signal indicated the association of the probe onto the target. We measured the mean lifetimes of both the high

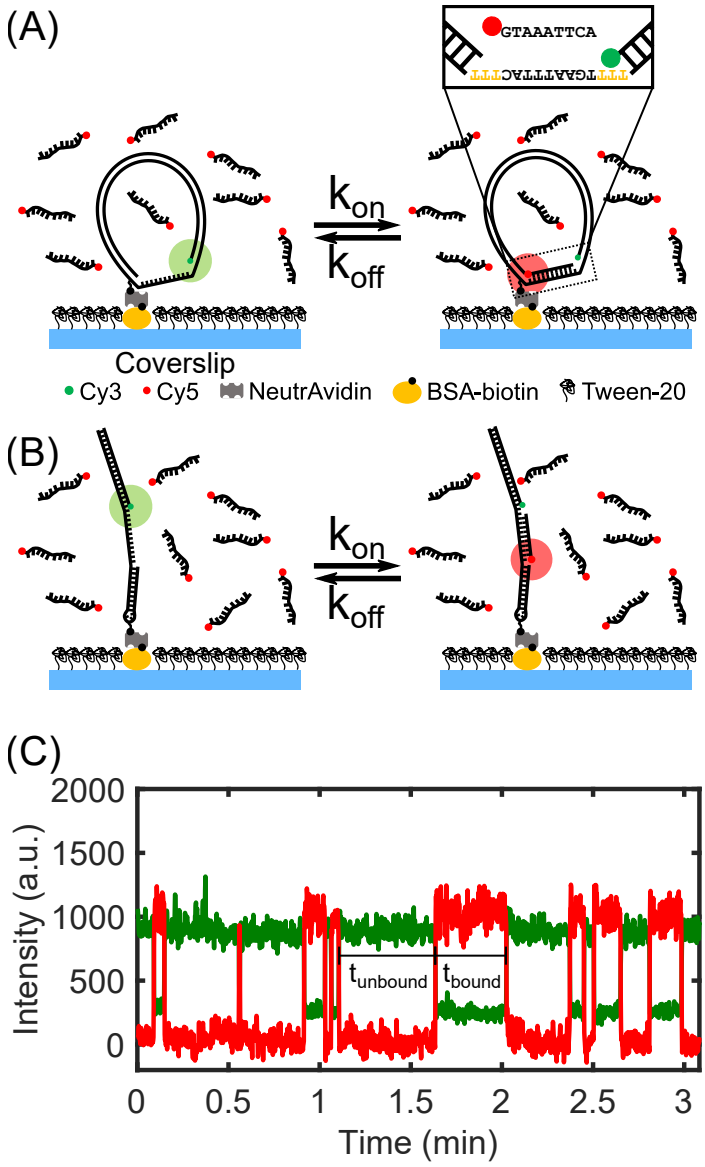


Figure 5.3: Experimental scheme. **(A)** Cy3-labeld cyclized DNA force generators were immobilized on a DDS-Tween-20 surface. An imaging buffer containing Cy5-labeld ss-DNA probes complementary to the ssDNA target of the cyclized DNA was flowed into the flow channel. Duplex formation and dissocation kinetics between the probe and target were monitored by FRET. The zoomed-in image highlights the 9-nt ssDNA target sequence (TGAATTTAC, black), which are capped by 3-nt poly T linkers (yellow). **(B)** For control experiments, partially duplexed, linear constructs were used instead of cyclized constructs. **(C)** Representative time trajectories from a cylized DNA force generator. Cy5 and Cy3 intensity traces are shown in red and green, respectively. The Cy5 signal increases when a DNA probe binds to the immobilized DNA. The high-Cy5 and high-Cy3 states are referred to as the bound and unbound states, respectively.

(i.e., high Cy5) and low FRET (i.e., high Cy3) states for approximately 100 immobilized molecules. This FRET measurement was repeated two to three times for each of the DNA constructs we prepared.

5.1.3 Preliminary results and discussion

Force-dependent transitions between the duplexed and separated states of short DNA are generally understood in terms of a diffusion process along a one-dimensional free energy landscape. The x-axis of the free energy landscape, or the reaction coordinate, is often the distance between the two points being pulled apart, in the case of a force-driven separation process. Under this assumption, the duplexed and separated states under a stretching force are well-defined as the local free energy minima separated by a barrier indicating the transition state. As the stretching force increases, the transition state energy barrier will be lowered, making the diffusive crossing from the bound state to the unbound state easier. The characteristic rate of reaching the transition state from the bound state, k_{off}, is related to the stretching force, F, as follows [53]:

$$k_{off}(F) = k_0(0) \exp\left(\frac{\Delta G^{\ddagger}(F)}{k_B T}\right) \tag{5.1}$$

where k_0 is the attempt frequency and $\Delta G^{\ddagger}(F)$ is the force-dependent height of the energy barrier. In Bell's model, $\Delta G^{\ddagger}(F)$ is simply modeled as $F\Delta x^{\ddagger}$, where $\Delta x^{\ddagger}$ is the distance from the bound state to the transition state in terms of the stretching reaction coordinate. Hence, the log of k_{off} is a linear function of F. In many cases, this relationship could explain the force-driven separation kinetics fairly well [159]. Its assumption is not quite realistic, however, given that $\Delta x^{\ddagger}$ could also depend on the force. Considering the force of

dependence of $\Delta x^{\ddagger}$ following Ref. [73, 139], we have

$$\begin{aligned} \Delta G^{\ddagger}(F) &= \Delta G^{\ddagger}(0) - l \int_0^F \Delta x^{\ddagger}(f) df \\ &= \Delta G^{\ddagger}(0) - l \int_0^F (x^{\ddagger}(f) - x_b(f)) df \end{aligned} \tag{5.2}$$

where l is the intrinsic contour length of the molecule being pulled, and $x^{\ddagger}(f)$ and $x_b(f)$ are extensions of the transition state and bound state configurations, respectively, at a constant force, f. In Whitley et al.'s study [73], $x_b(f)$ and $x^{\ddagger}(f)$ were found based on the force-extension relation of WLC. Whereas $x_b(f)$ was modeled by the well-known double-stranded DNA force-extension relation, the force-extension relation of the transition state DNA, $x^{\ddagger}(f)$ was less well-characterized. Nonetheless, modeling $x^{\ddagger}(f)$ as the WLC model with a persistence length (l_p) equal to $\sim$2.6 nm and the contour lengths per base pair (h) equal to 0.54 nm/nt was enough to capture the non-exponential behavior of $k_{off}(F)$ in their study. The key observation from their model is that it predicts that low forces could stabilize the duplex, resulting in a roll-over behavior in $k_{off}(F)$ below $F \sim 5$ pN.

To probe the force-dependent kinetics of short DNA duplex formation and dissociation at the critical force range below 5 pN, we performed a similar set of experiments as in Whitley et al.[73]. In our experiment, we generated low stretching forces across a short ssDNA by using a set of DNA-based force generators constructed from cyclized DNA molecules of sizes equal to 158 bp, 126 bp, and 84 bp. We expected cycled DNA at these lengths to generate pN-range stretching forces. The change in duplex length would allow us to probe the low-force dependence.

In Fig 5.4(A), we show the mean bound and unbound state lifetimes (τ_{bound} and $\tau_{unbound}$). We found that the bound state lifetime was dependent on DNA length. The bound state lifetime decreased as the length of the DNA construct decreased. The decrease in τ_{bound} is expected since the outward stretching force would be greater with shorter lengths as they require sharper bending. As a control, we measured τ_{bound} in the absence of the stretching

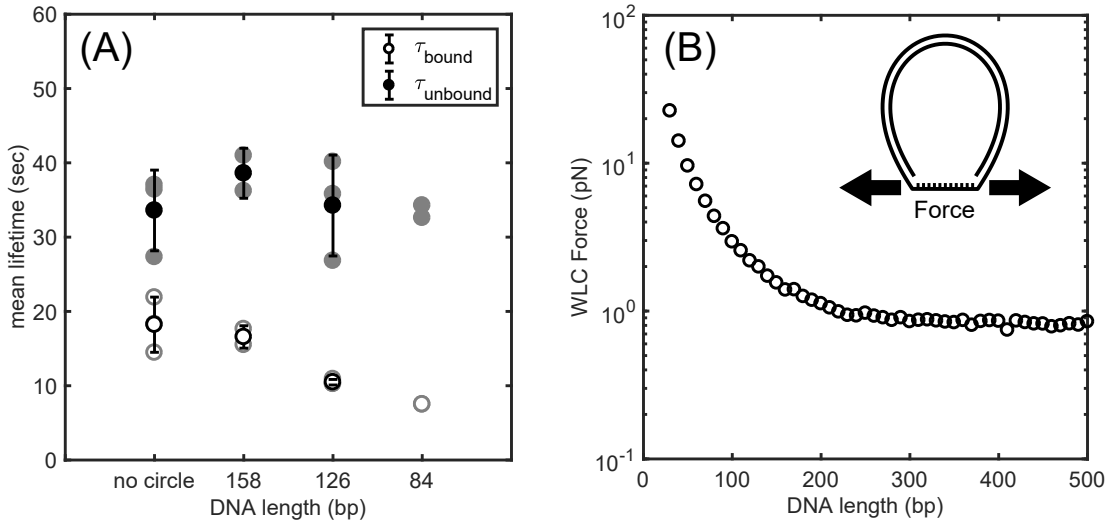


Figure 5.4: **(A)** Mean bound (open circles) and unbound state (closed circles) lifetimes as a function DNA length. Results from the control DNA construct with no DNA loop are also plotted at the "no circle" tick mark. The error bars represent repeatability from three trials, with black symbols representing the mean. Grey symbols represent individual trials. **(B)** Stretching force predicted by the WLC model as a function of DNA length. Schematic at the top right-hand corner represents the outward stretching force across the ssDNA target sequence exerted by the duplexed loop.

force by using partially duplexed linear DNA. From this measurement, we found that the bound state lifetime to be the longest, in accordance with our expectation.

On the contrary, the unbound state lifetime ($\tau_{unbound}$) did not depend on the length of cyclized DNA. In fact, the force-dependence of the association rate was not as prominent as the case of dissociation as shown by Whitley et al.[73]. In agreement with this observation, we found that the unbound state lifetimes from the cyclized constructs were quite comparable to that of the linear DNA control.

Next, we estimated the stretching force by cyclized DNA based on the WLC model. For this estimation, we performed the biased Monte Carlo simulations of coarse-grained DNA as described in Chapter 2. Here, we chose the end-to-end distance between the ground-state cyclized DNA to be 7 nm, which would equal to the contour length of the ssDNA target with a bound probe (i.e., 2 $\times$ 3 nt + 9 bp). As shown in Fig 5.4(B), the stretching

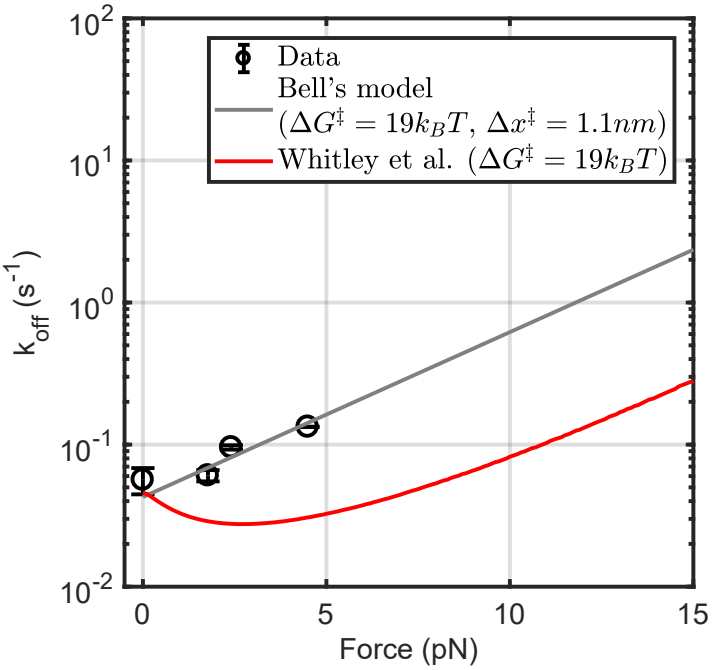


Figure 5.5: Data from Figure 5.4(A) are plotted in terms of dissociation rate, k_{off}, as a function WLC force. The data points are fitted against two different models, namely, Bell's model (grey line) and Whitley et al.'s roll-over model [73]. The fitting parameters in each model are shown in the legend.

force increases with decreasing DNA length. At lengths below ~150 bp, the stretching force increases even more steeply, since DNA below 150 bp behaves as an elastic beam. Note that the stretching force does not decay to zero at large lengths since force can be generated by the entropic fluctuations of very long DNA. This entropic force appears to plateau around 1 pN. From the force simulations, we find the three different DNA lengths we tried, 158 bp, 126 bp, and 84 bp, would correspond to 1.8 pN, 2.4 pN, and 4.5 pN in forces, respectively.

In Fig 5.5, we plotted the dissociation rate, k_{off} (=$1/\tau_{bound}$), from our data as a function of force. This plot allows us to directly compare our data against Whitley et al.'s force dependent dissociation model (or the "roll-over" model) [73]. Following this model, we re-produced the theoretical force-dependence of k_{off}. To fit the model to our data, we used

the data point from the zero-force control construct to estimate $\Delta G^{\dagger}(0)$, which we found $19k_BT$ as the best fitting value. From the fitting, we find that this model fits poorly with our data points, although there are not many data points. Nevertheless, because the data points are fairly sparsely distributed in the range where the roll-over is the most noticeable (between 0 pN and 5 pN), we believe these points are enough to provide a clue as to whether or not the force-dependence is exponential. Instead of a roll-over, we see in our data a monotonic force-dependence, which could possibly be fitted by Bell's model (grey line in Figure 5.5). For Bell's model prediction, we used the same value of $\Delta G^{\dagger}(0)$ as before and used 1.1 nm as $x^{\dagger}$, which would capture the high-force behavior of the roll-over model. It is important to point out that Whitley et al. also posited a possibility that their roll-over model could be more inaccurate towards zero force, as the one-dimensional end-to-end extension coordinate could be inappropriate for describing annealing and dissociation transitions at small forces. The zero-force data point from Ref. [151] plotted along with the 9-bp DNA probe in Figure 2A of their study was provided to support this scenario. Additionally, we find that the proposed strong roll-over was barely noticeable in their data. Only one or two DNA probes under different experimental conditions, out of 14 different experiments (i.e., four DNA probes at 100 mM NaCl, five DNA probes at 100 mM NaCl + 20 mM $MgCl_2$, and five DNA–RNA hybrids at 100 mM NaCl + 20 mM $MgCl_2$), appeared to support the roll-over behavior.

The monotonic dependence without roll-over exhibited in our data suggests that Bell's model could still be fairly accurate near the low force regime. Although the number of data points is low, the fact that there is no rise in our data at zero-force appears meaningful enough to suggest reconsideration of the roll-over model. A previous study conducted by our lab [33] also showed that the force-dependent dissociation of short DNA at a low force regime agreed well with Bell's model with no roll-over, although the pulling geometry was slightly different from that in the present study. Alternatively, rather than rolling-over at low forces below $\sim$2 pN, our data could also be showing plateauing of the dissociation rate

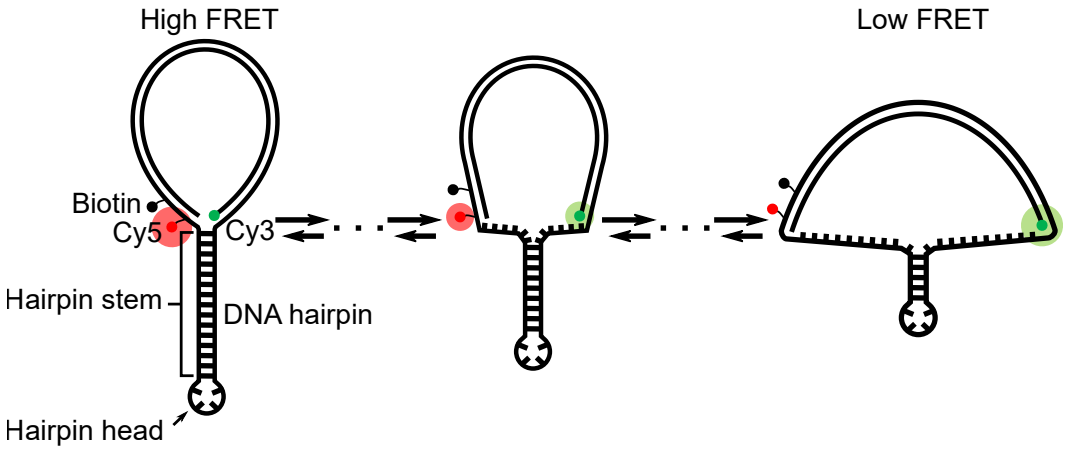


Figure 5.6: Schematic of DNA hairpin unzipping by cyclized DNA. A DNA hairpin is linked to a cyclized DNA molecule. The stem of the hairpin is long enough to be stably duplexed well above the room temperature. The length of the cyclized DNA should be less than the persistence length of DNA to act as a force generator. The 5′ and 3′ ends of DNA hairpin are pulled apart due to the elasticity of the cyclized DNA, thus; unpairing would be more preferred than annealing. This process unzips base pairs from the end of the hairpin stem towards the head. The high FRET signal disappears when the unpaired strands of the stem are stretched long enough beyond the working distance of FRET. The experiment will focus on collecting the probability distribution of hairpin lifetimes (i.e. duration between the highest FRET and the lowest FRET states).

[89]. Collecting more data outside and in between the force values tested in the present study would be very useful. Cyclized DNA approximately 70 bp in length appears to be the smallest size that could be made without kinking, generating an stretching force comparable to 10 pN. Cyclized DNA around 200 bp, where it reaches to the entropic force limit, would provide critical data points to confirm whether the force verses rate curve shows a roll-over, plateau, or linear dependence.

5.2 Measuring the unzipping kinetics of DNA hairpins

The cyclized DNA force generator can also be used to measure the force-induced unzipping kinetics of a DNA hairpin. The idea here is to replace the ssDNA target sequence in the cyclized DNA construct described in the previous section with a self-complementary DNA hairpin as shown in Figure 5.6. Under this design, the 5′ and 3′ ends of the hairpin structure

would be pulled apart by the cyclized duplex. In this case, a FRET donor–acceptor pair should be labeled near the top of the hairpin stem, such that the high FRET signal is only obtained when the construct is in the zipped state. Although there have been numerous hairpin unzipping studies [86, 87], most have focused on quantifying either the mean rupture time as a function of force or the mean rupture force at which the rupture probability is 0.5. Instead of probing such quantities, we shall focus on the survival probability of the hairpin as a function of time, under a stretching force. Assuming the unzipping process is a series of energy-barrier crossings, we hypothesize that the survival probability will not be a perfectly exponential function, which would represent the process of a single dominant energy barrier, but will rather be somewhat like a lagged exponential or a gamma distribution function. While this hypothesis has been tested theoretically[135, 136], currently no experiment has done in this context. High-throughput collection of unzipping time distribution appears very feasible with the FRET-based cyclized DNA force generator assay. Alternatively, the hairpin loop design from Chapter 4 could be used for this experiment in principle, provided the hairpin sticky ends are long enough to stabilize the hairpin stem fairly efficiently.

5.3 Sequence-dependent flexibility of single mismatches

From our study in Chapter 4, we have learned that DNA cyclization could be applied to study the mechanical properties of mismatched DNA, although great caution should be made concerning the choice of sticky-end pairing geometry. For probing the bending elasticity of DNA with base pair mismatch, the hairpin sticky pair should be used to avoid cooperative kinking. As an extension of the mismatch experiment described in Chapter 4, the hairpin cyclization assay could be used to address questions regarding the sequence-dependent effect of mismatched base pairs.

DNA flexibility has been shown to be sequence-dependent [97]. Depending on how a DNA sequence is arranged, the global bendability of DNA can change. Similarly, for

mismatched base pairs, local mechanical properties such as shape and flexibility will vary depending on the bases in the mismatched pair and also on the neighboring base pairs. Sequence-dependent mismatch flexibility may provide insights into how structural distortions influences protein-DNA interactions [161, 162, 163] and how mismatches are recognized and repaired by mimsatch repair enzymes [134, 164]. While some experimental studies have investigated the flexibility of mismatched DNA [57, 130, 142], not many have focused on flexibility in the context of sequence dependence.

Perhaps the most relevant experimental work is a study by Fields et al. [51] in which bendedness of DNA with a central mismatch of various sequences was measured using a method combining gel-electrophoresis and FRET. In their study, the bendability of base pair mismatch was studied using DNA-based molecular vises. Similar to our hairpin loop from Chapter 4, this molecular vise consisted of a DNA duplex surrounded by sticky ends, which could anneal like the stem of a hairpin. In this construct, a FRET pair was labeled near the fork side of the stem, such that the high FRET signal was only possible when the construct was fully zipped. Based on the FRET efficiency as a function of the intervening duplex size, the bendability of the duplex was inferred. Fields and colleagues tested eight different central base pair mismatches and found that not all mismatches had the same bendability, with some sequences being stiffer than others. They also found that sequence-dependent bendability of the mismatches was correlated to their thermodynamic stability to some degree.

While Fields et al.'s mismatch study [51] provided some insights regarding the sequence-dependent flexibility of mismatched DNA, their method was somewhat limited in that the measured flexibility was only valid under strong bending conditions; that is, the intervening loop in a molecular vise only represents fluctuations with high curvature. In this regard, our hairpin loop assaying appears more versatile for probing mismatch flexibility, as our assay can separately measure the looping and unlooping processes, which represent both low- and high-curvature behaviors.

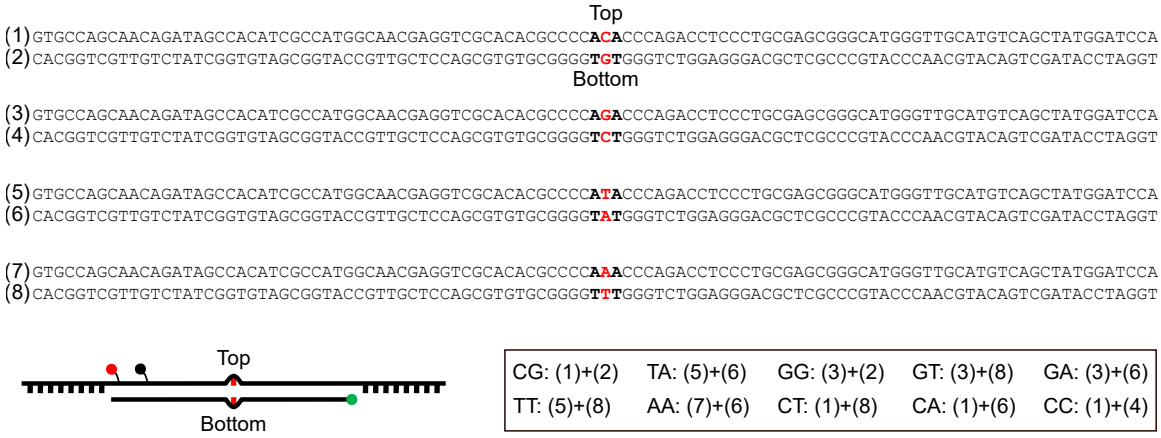


Figure 5.7: Four DNA sequences prepared in this study. The sequence of these DNA are identical except for the central base pair the middle colored in red. The nearest-neighbors of the central base pair are shown as bold letters. The top and bottom strand of each DNA are numbered. The odd-numbered strands indicate the top strand that contains hairpin sticky ends, Cy5, and biotin. The even-numbered strands indicate the bottom strand with Cy3. The figure in the bottom left shows a schematic of the final hairpin molecule as a result of a strand exchange reaction of different strand combinations. The red, black, and green dots represent Cy5, biotin, and Cy3 respectively. The central mismatched base pair is represented as red bases inside a bulge. The box in the bottom right shows the 8 different mismatched DNA and 2 intact DNA molecules made in this study. These sequences are named according to the central red bases in their top and bottom strands. Each name specifies what DNA strands are used in a strand exchange reaction. For example, the two intact DNA molecules, CG and TA are made from mixing Strand (1) with Strand (2), and Strand (5) with Strand (6).

5.3.1 Materials and methods

The DNA preparation step is mostly identical to that described in Chapter 4. Briefly, we first prepared a set of four DNA sequences differ among each other only in one base pair in the middle, but otherwise the same (Figure 5.7). These sequences were directly synthesized as single-stranded oligonucleotides, duplexed by PCR, and cloned into the *E. coli* cells as described before. The cloned sequences were extracted by colony-PCR and the quality of these sequences were confirmed by sequencing. The four DNA sequences were, then, modified to incorporate Cy5, Biotin, and sticky ends by PCR reactions. Using the same four DNA sequences in separate PCR reactions, we prepared another batch of identical DNA sequences labeled with Cy3. By mixing various Cy5-labeld sequences with Cy3-labeled

sequences in strand-exchange reactions, we generated 8 different base pair mismatches (see Fig 5.7). In addition to the mismatched DNA, we also prepared two intact sequences. The sequence of sticky ends are identical as the sticky ends of the hairpin DNA construct found in Chapter 4. The ten different DNA sequences (8 mismatches + 2 intact) are named according to the varying middle base pair "XY", where X and Y are the varying bases of the Cy5 and Cy3 strands, respectively.

5.3.2 Preliminary results and future outlook

Here, we present some preliminary results using the hairpin loop construct of 105 bp in length. We measured the unlooping rates (k_{unloop}) of various single mismatches by placing them near the center of the hairpin loop following the buffer-exchange method. The molecules were incubated in an imaging buffer containing 2 M NaCl inside a flow-cell and unlooping was induced by flowing in a new imaging buffer with 150 mM NaCl.

The unlooping rates of the mismatched molecules are plotted in Figure 5.8(A) along with the two intact molecules. We expected that the intact DNA molecules to be the fastest unlooping DNA as base pair mismatch could cause structural distortion in DNA and act as a flexible hinge, relieving the bending stress. We, first noticed that the unlooping rate exhibited some mismatch sequence-dependence. Most notably, the GT showed more than two-fold faster unlooping than the CC mismatch. While the unlooping rates of mismatched molecules were mostly slower than the intact molecule, the GG and GT mismatches appeared to induce faster unlooping than the intact. Intuitively, faster unlooping rates suggest higher stiffness. According to this assumption, DNA with GG and GT mismatches could be more rigid than the conventional B-DNA. The unusually high stiffness of these mismatches could be related to their unusually high structural stability [56]. As indicated by Figure 5.8(B), GG and GT are generally less destabilizing than other mismatches and depending of the sequence context, GG and GT could even stabilize the duplex.

As shown by Fields et al.[51], the sequence-dependent mismatch flexibility might be

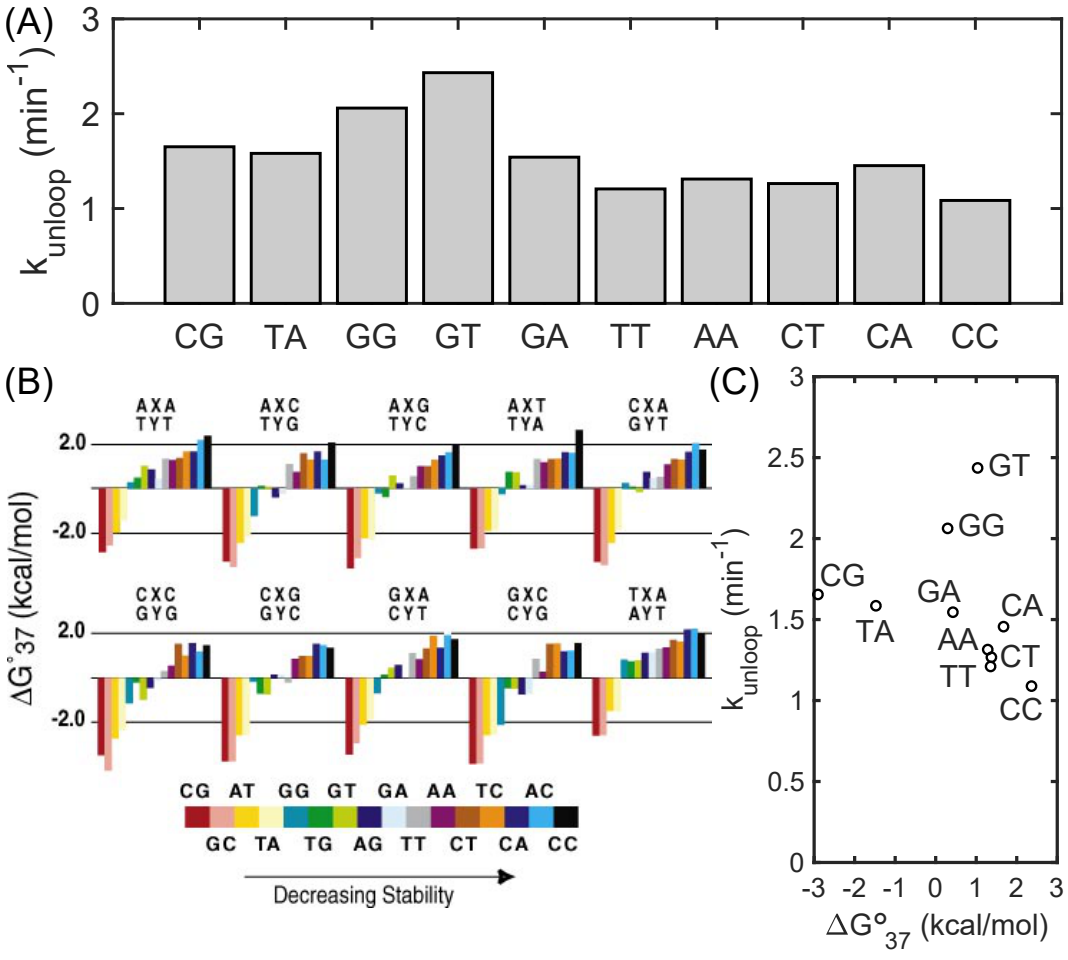


Figure 5.8: **(A)** Sequence-dependent unlooping rates of the 10 DNA sequences (2 intact + 8 mismatches). **(B)** Free energies of single mismatches in the context of different nearest-neighbor sequences. This figure is redrawn from Figure 4 in reference [137] with permission from Annual Reviews, Inc.. **(C)** Correlation between the unlooping rates and the thermal stability of the tested DNA sequences. (Spearman correlation: -0.8)

related to the thermodynamic stability of mismatch. We, thus, calculated free energies of the tested eight mismatches as well, considering the two AT pairs as the neighboring base pairs. The calculated free energies are plotted in Figure 5.8(B) in the order of their average thermal stability over all possible nearest-neighbors. Based on these values, we found that the unlooping rates of the tested mismatches were correlated with their free energy to some extent (Figure 5.8(C)), with Spearman correlation of -0.8, in accordance with Fields et al.'s findings.

The looping kinetics, which have not yet been tested, can also be similarly measured

using the buffer-exchange method, but with the reversed salt condition. Alternatively, the experiment can be conducted non-perturbatively without buffer exchange. In such a case, reversible looping and unlooping transitions could be monitored under one intermediate salt condition without buffer exchange. It will be interesting to see how the looping rates of these mismatches are correlated to their thermal stability. The looping experiment would also allow us to investigate how the flexibility of base pair mismatch vary under different curvature constraints. The above unlooping experiments as well as with Fields et al.'s study[51] only provide the sequence-dependent bendability under sharp curvature constraints because experimental measurable reflects fluctuations within the looped configurations. On the other hand, the looping dynamics may be different from unlooping since it is free of such constraints. The key advantage of the hairpin looping assay is that we can separately measure and compare the looping and unlooping rates. This experiment would certainly provide further insights concerning the mechanics of DNA in the presence of base pair mismatch.

Appendices

APPENDIX A

SUPPLEMENTARY MATERIALS FOR CHAPTER 3

A.1 DNA sequences

All molecules in DNA sets 1 and 2 include the common adapter sequences (20 bp) at both ends, which are also present in the PCR primer pairs. PCR primers and blocking oligos are hybridized to each other to make the partial duplexes that are sticky on one end and blunt on the other: Blocking-Cy5full and Blocking-Cy5gap are hybridized with the forward primers of full and gapped sticky ends, and Blocking-Cy3 hybridizes with the backward primers of both full and gapped sticky ends, respectively. These partial duplexes are used to measure the association rate (k_{on}) between the sticky ends and the lifetime (τ_{on}) of the linker duplex. Below tables are the list of DNA sequences, PCR primers, and blocking oligonucleotides.

DNA set 1 (5′ to 3′)	
94 bp	GTGCCAGCAACAGATAGCCACCGGAGCCACACCGGTGCAAACCTC AGCAAGCAGGGTGTGGAAGTAGGACATTTCCCATTCGAGCTCGTT GTAG
96 bp	GTGCCAGCAACAGATAGCCATCCGGAGCCACACCGGTGCAAACCT CAGCAAGCAGGGTGTGGAAGTAGGACATTTTCCCATTCGAGCTCG TTGTAG
98 bp	GTGCCAGCAACAGATAGCCATTCCGGAGCCACACCGGTGCAAACC TCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCCCCATTCGAGCT CGTTGTAG
	Continued on next page

DNA set 1 (5′ to 3′) – continued from previous page	
100 bp	GTGCCAGCAACAGATAGCCACTTCCGGAGCCACACCGGTGCAAAC CTCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCACCCATTCGAG CTCGTTGTAG
102 bp	GTGCCAGCAACAGATAGCCAACTTCCGGAGCCACACCGGTGCAAA CCTCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCATCCCATTCG AGCTCGTTGTAG
104 bp	GTGCCAGCAACAGATAGCCAAACTTCCGGAGCCACACCGGTGCAA ACCTCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCATGCCCATT CGAGCTCGTTGTAG
106 bp	GTGCCAGCAACAGATAGCCATAACTTCCGGAGCCACACCGGTGCA AACCTCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCATGTCCCA TTCGAGCTCGTTGTAG
108 bp	GTGCCAGCAACAGATAGCCATTAACTTCCGGAGCCACACCGGTGC AAACCTCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCATGTCCC CATTCGAGCTCGTTGTAG
110 bp	GTGCCAGCAACAGATAGCCAGTTAACTTCCGGAGCCACACCGGTG CAAACCTCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCATGTCA CCCATTCGAGCTCGTTGTAG
112 bp	GTGCCAGCAACAGATAGCCACGTTAACTTCCGGAGCCACACCGGT GCAAACCTCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCATGTC AGCCCATTCGAGCTCGTTGTAG
114 bp	GTGCCAGCAACAGATAGCCAGCGTTAACTTCCGGAGCCACACCGG TGCAAACCTCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCATGT CAGGCCCATTCGAGCTCGTTGTAG
Continued on next page	

DNA set 1 (5′ to 3′) – continued from previous page	
116 bp	GTGCCAGCAACAGATAGCCAAGCGTTAACTTCCGGAGCCACACCG GTGCAAACCTCAGCAAGCAGGGTGTGGAAGTAGGACATTTTCATG TCAGGCCCCATTCGAGCTCGTTGTAG

Table A.1: List of DNA set 1.

DNA set 2 (5′ to 3′)	
94 bp	GTGCCAGCAACAGATAGCCACATGGCAACGAGGTCGCACACGCCCCACACCCAGACCTCCCTGCGAGCGGGCATCCCATTCGAGCTCGTTGTTG
108 bp	GTGCCAGCAACAGATAGCCACGATCGCCATGGCAACGAGGTCGCACACGCCCCACACCCAGACCTCCCTGCGAGCGGGCATGGGTACACCCATTCGAGCTCGTTGTAG
110 bp	GTGCCAGCAACAGATAGCCAGCGATCGCCATGGCAACGAGGTCGCACACGCCCCACACCCAGACCTCCCTGCGAGCGGGCATGGGTACAACCCATTCGAGCTCGTTGTAG
112 bp	GTGCCAGCAACAGATAGCCACGCGATCGCCATGGCAACGAGGTCGCACACGCCCCACACCCAGACCTCCCTGCGAGCGGGCATGGGTACAATCCCATTCGAGCTCGTTGTAG
114 bp	GTGCCAGCAACAGATAGCCAGCGCGATCGCCATGGCAACGAGGTCGCACACGCCCCACACCCAGACCTCCCTGCGAGCGGGCATGGGTACAATGCCCATTCGAGCTCGTTGTAG
116 bp	GTGCCAGCAACAGATAGCCACGCGCGATCGCCATGGCAACGAGGTCGCACACGCCCCACACCCAGACCTCCCTGCGAGCGGGCATGGGTACAATGTCCCATTCGAGCTCGTTGTAG
118 bp	GTGCCAGCAACAGATAGCCAACGCGCGATCGCCATGGCAACGAGGTCGCACACGCCCCACACCCAGACCTCCCTGCGAGCGGGCATGGGTACAATGTCCCCATTCGAGCTCGTTGTAG
120 bp	GTGCCAGCAACAGATAGCCACACGCGCGATCGCCATGGCAACGAGGTCGCACACGCCCCACACCCAGACCTCCCTGCGAGCGGGCATGGGTACAATGTCCCCCATTCGAGCTCGTTGTAG

Continued on next page

DNA set 2 (5′ to 3′) – continued from previous page	
122 bp	GTGCCAGCAACAGATAGCCACCACGCGCGATCGCCATGGCAACGA GGTCGCACACGCCCCACACCCAGACCTCCCTGCGAGCGGGCATGG GTACAATGTCCCCCCATTCGAGCTCGTTGTAG
124 bp	GTGCCAGCAACAGATAGCCACCCACGCGCGATCGCCATGGCAACG AGGTCGCACACGCCCCACACCCAGACCTCCCTGCGAGCGGGCATG GGTACAATGTCCCCCCCATTCGAGCTCGTTGTAG
153 bp	GTGCCAGCAACAGATAGCCAGGGGAAAGACCACACCCACGCGCG ATCGCCATGGCAACGAGGTCGCACACGCCCCACACCCAGACCTCC CTGCGAGCGGGCATGGGTACAATGTCCCCGTTGCCACAGAGACCC CCATTCGAGCTCGTTGTAG
189 bp	GTGCCAGCAACAGATAGCCACAGCAACGGGCAACCGTTTGGGGAA AGACCACACCCACGCGCGATCGCCATGGCAACGAGGTCGCACACG CCCCACACCCAGACCTCCCTGCGAGCGGGCATGGGTACAATGTCC CCGTTGCCACAGAGACCACTTCGTAGCACAGCGCCCCATTCGAGC TCGTTGTAG

Table A.2: List of DNA set 2.

Full sticky ends PCR primer pairs (5′ to 3′)	
Forward	TGAATTTACG[Cy5dT]GCCAGCAACAGA[BiotindT]AGC
Backward	GTAAATTCAC[Cy3dT]ACAACGAGCTCGAATGGG
Gapped sticky ends PCR primer pairs (5′ to 3′)	
Forward	TGAATTTACGCTG[Cy5dT]GCCAGCAACAGA[BiotindT]AGC CA
Backward	[Cy3]GTAAATTCACGACTACAACGAGCTCGAATGGG
Blocking oligos for making partial duplexes (5′ to 3′)	
Blocking-Cy5full	GCTATCTGTTGCTGGCAC
Blocking-Cy5gap	TGGCTATCTGTTGCTGGCAC
Blocking-Cy3	CCCATTCGAGCTCGTTGTAG

Table A.3: List of PCR primers and blocking oligonucleotides.

A.2 Free energy difference between the singly-kinked and doubly-kinked loops

Replacing $J_{\phi\theta}$ with the J factor estimated with an end-to-end distance equal to 3 nm (i.e. the contour length of the linker duplex) in Equation (2) in the main text, we can calculate the free energy difference between the singly and doubly-kinked loops. Taking into account ΔG_{ST} ($\approx$ 3.5 k_BT), we find that the doubly-kinked loop is thermodynamically less favorable than the singly-kinked loop by $\sim$ 2 k_BT, which corresponds to an equilibrium fraction of 0.14, in the range between 95 bp and 125 bp. Therefore, we focus on our discussion on transitions between the teardrop state with a single open nick and the smooth state.

REFERENCES

[1] A. Travers and J. Thompson, "An introduction to the mechanics of dna," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 362, no. 1820, pp. 1265–1279, 2004.

[2] Y.-J. Chen, S. Johnson, P. Mulligan, A. J. Spakowitz, and R. Phillips, "Modulation of DNA loop lifetimes by the free energy of loop formation," *Proceedings of the National Academy of Sciences*, vol. 111, no. 49, pp. 17 396–17 401, 2014.

[3] S. Johnson, M. Lindén, and R. Phillips, "Sequence dependence of transcription factor-mediated DNA looping," *Nucleic Acids Research*, vol. 40, no. 16, pp. 7728–7738, 2012.

[4] S. E. Halford, A. J. Welsh, and M. D. Szczelkun, "Enzyme-mediated DNA looping," *Annual Review of Biophysics and Biomolecular Structure*, vol. 33, no. 1, pp. 1–24, 2004.

[5] G. R. Bellomy, M. C. Mossing, and M. T. Record, "Physical properties of DNA in vivo as probed by the length dependence of the lac operator looping process," *Biochemistry*, vol. 27, no. 11, pp. 3900–3906, 1988.

[6] J. Müller, S. Oehler, and B. Müller-Hill, "Repression oflacPromoter as a function of distance, phase and quality of an AuxiliarylacOperator," *Journal of Molecular Biology*, vol. 257, no. 1, pp. 21–29, 1996.

[7] N. A. Becker, J. D. Kahn, and L. J. Maher, "Bacterial repression loops require enhanced DNA flexibility," *Journal of Molecular Biology*, vol. 349, no. 4, pp. 716–730, 2005.

[8] A. J. Andrews and K. Luger, "Nucleosome structure (s) and stability: Variations on a theme," *Annual review of biophysics*, vol. 40, pp. 99–117, 2011.

[9] C. Bustamante, Y. R. Chemla, N. R. Forde, and D. Izhaky, "Mechanical processes in biochemistry," *Annual review of biochemistry*, vol. 73, no. 1, pp. 705–748, 2004.

[10] R. Schleif, "DNA looping," *Annual review of biochemistry*, vol. 61, no. 1, pp. 199–223, 1992.

[11] D. Shore, J. Langowski, and R. L. Baldwin, "DNA flexibility studied by covalent closure of short fragments into circles," *Proceedings of the National Academy of Sciences*, vol. 78, no. 8, pp. 4833–4837, 1981.

[12] D. M. Crothers, J. Drak, J. D. Kahn, and S. D. Levene, *[1] DNA bending, flexibility, and helical repeat by cyclization kinetics*. Elsevier, 1992, pp. 3–29.

[13] J. P. Peters and L. J. Maher, "DNA curvature and flexibility in vitro and in vivo," *Quarterly Reviews of Biophysics*, vol. 43, no. 01, pp. 23–63, 2010.

[14] T. E. Cloutier and J. Widom, "Spontaneous sharp bending of double-stranded DNA," *Molecular Cell*, vol. 14, no. 3, pp. 355–362, 2004.

[15] R. Vafabakhsh and T. Ha, "Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization," *Science*, vol. 337, no. 6098, pp. 1097–1101, 2012.

[16] A. Vologodskii and M. D. Frank-Kamenetskii, "Strong bending of the DNA double helix," *Nucleic Acids Research*, vol. 41, no. 14, pp. 6785–6792, 2013.

[17] A. Vologodskii, Q. Du, and M. D. Frank-Kamenetskii, "Bending of short DNA helices," *Artificial DNA: PNA & XNA*, vol. 4, no. 1, pp. 1–3, 2013.

[18] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii, "Cyclization of short DNA fragments and bending fluctuations of the double helix," *Proceedings of the National Academy of Sciences*, vol. 102, no. 15, pp. 5397–5402, 2005.

[19] T. T. Le and H. D. Kim, "Studying DNA looping by single-molecule FRET," *Journal of Visualized Experiments*, no. 88, 2014.

[20] R. P. Bowater and Z. A. Waller, "DNA structure," in *eLS*. American Cancer Society, 2014, ISBN: 9780470015902.

[21] D. Shore and R. L. Baldwin, "Energetics of DNA twisting," *Journal of Molecular Biology*, vol. 170, pp. 957–981, 1983.

[22] W. H. Taylor and P. J. Hagerman, "Application of the method of phage t4 DNA ligase-catalyzed ring-closure to the study of DNA structure: Ii. nacl-dependence of DNA flexibility and helical repeat," *Journal of molecular biology*, vol. 212, no. 2, pp. 363–376, 1990.

[23] M. Vologodskaia and A. Vologodskii, "Contribution of the intrinsic curvature to measured DNA persistence length1," *Journal of molecular biology*, vol. 317, no. 2, pp. 205–213, 2002.

[24] A. Brunet, C. Tardin, L. Salome, P. Rousseau, N. Destainville, and M. Manghi, "Dependence of DNA persistence length on ionic strength of solutions with monova-

lent and divalent salts: A joint theory–experiment study," *Macromolecules*, vol. 48, no. 11, pp. 3641–3652, 2015.

[25] J. F. Marko, "Biophysics of protein–DNA interactions and chromosome organization," *Physica A: Statistical Mechanics and its Applications*, vol. 418, pp. 126–153, 2015.

[26] N. Becker, A. Rosa, and R. Everaers, "The radial distribution function of worm-like chains," *The European Physical Journal E*, vol. 32, no. 1, pp. 53–69, 2010.

[27] J. Shimada and H. Yamakawa, "Ring-closure probabilities for twisted wormlike chains. application to dna," *Macromolecules*, vol. 17, no. 4, pp. 689–698, 1984.

[28] C. Hyeon and D. Thirumalai, "Kinetics of interior loop formation in semiflexible chains," *The Journal of chemical physics*, vol. 124, no. 10, p. 104 905, 2006.

[29] J.-F. Allemand, S. Cocco, N. Douarche, and G. Lia, "Loops in DNA: An overview of experimental and theoretical approaches," *The European Physical Journal E*, vol. 19, no. 3, pp. 293–302, 2006.

[30] L. Czapla, D. Swigon, and W. K. Olson, "Sequence-dependent effects in the cyclization of short dna," *Journal of Chemical Theory and Computation*, vol. 2, no. 3, pp. 685–695, 2006.

[31] N. J. Agrawal, R. Radhakrishnan, and P. K. Purohit, "Geometry of mediating protein affects the probability of loop formation in dna," *Biophysical journal*, vol. 94, no. 8, pp. 3150–3158, 2008.

[32] H. Salari, B. Eslami-Mossallam, S. Naderi, and M. Ejtehadi, "Extreme bendability of dna double helix due to bending asymmetry," *The Journal of chemical physics*, vol. 143, no. 10, 09B608_1, 2015.

[33] T. T. Le and H. D. Kim, "Probing the elastic limit of DNA bending," *Nucleic Acids Research*, vol. 42, no. 16, pp. 10 786–10 794, 2014.

[34] R. A. Forties, R. Bundschuh, and M. G. Poirier, "The flexibility of locally melted dna," *Nucleic acids research*, vol. 37, no. 14, pp. 4580–4586, 2009.

[35] T. E. Cloutier and J. Widom, "DNA twisting flexibility and the formation of sharply looped protein-DNA complexes," *Proceedings of the National Academy of Sciences*, vol. 102, pp. 3645–3650, 2005.

[36] J. Yan and J. F. Marko, "Localized single-stranded bubble mechanism for cyclization of short double helix dna," *Physical review letters*, vol. 93, no. 10, p. 108 108, 2004.

[37] P. A. Wiggins, R. Phillips, and P. C. Nelson, "Exact theory of kinkable elastic polymers," *Physical Review E*, vol. 71, no. 2, p. 021 909, 2005.

[38] G. Kaur, J. S. Lewis, and A. M. van Oijen, "Shining a spotlight on dna: Single-molecule methods to visualise dna," *Molecules*, vol. 24, no. 3, p. 491, 2019.

[39] P. Ranjith, P. S. Kumar, and G. I. Menon, "Distribution functions, loop formation probabilities, and force-extension relations in a model for short double-stranded DNA molecules," *Physical review letters*, vol. 94, no. 13, p. 138 102, 2005.

[40] F. Lankaš, R. Lavery, and J. H. Maddocks, "Kinking occurs during molecular dynamics simulations of small DNA minicircles," *Structure*, vol. 14, no. 10, pp. 1527–1534, 2006.

[41] J. Mitchell, C. Laughton, and S. A. Harris, "Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled dna," *Nucleic acids research*, gkq1312, 2011.

[42] R. N. Irobalieva, J. M. Fogg, D. J. Catanese, T. Sutthibutpong, M. Chen, A. K. Barker, S. J. Ludtke, S. A. Harris, M. F. Schmid, W. Chiu, *et al.*, "Structural diversity of supercoiled dna," *Nature communications*, vol. 6, 2015.

[43] P. J. Mulligan, Y.-J. Chen, R. Phillips, and A. J. Spakowitz, "Interplay of protein binding interactions, DNA mechanics, and entropy in DNA looping kinetics," *Biophysical Journal*, vol. 109, no. 3, pp. 618–629, 2015.

[44] C. Kim, O.-c. Lee, J.-Y. Kim, W. Sung, N. K. Lee, *et al.*, "Dynamic release of bending stress in short dsDNA by formation of a kink and forks," *Angewandte Chemie International Edition*, vol. 54, no. 31, pp. 8943–8947, 2015.

[45] P. Cong, L. Dai, H. Chen, J. R. van der Maarel, P. S. Doyle, and J. Yan, "Revisiting the anomalous bending elasticity of sharply bent dna," *Biophysical journal*, vol. 109, no. 11, pp. 2338–2351, 2015.

[46] L. Han, H. G. Garcia, S. Blumberg, K. B. Towles, J. F. Beausang, P. C. Nelson, and R. Phillips, "Concentration and length dependence of DNA looping in transcriptional regulation," *PloS one*, vol. 4, no. 5, e5621–e5621, 2009.

[47] D. G. Priest, L. Cui, S. Kumar, D. D. Dunlap, I. B. Dodd, and K. E. Shearwin, "Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the lac and λ repressors," *Proceedings of the National Academy of Sciences*, vol. 111, no. 1, pp. 349–354, 2014.

[48] H. Jacobson and W. H. Stockmayer, "Intramolecular reaction in polycondensations. i. the theory of linear systems," *The Journal of Chemical Physics*, vol. 18, no. 12, pp. 1600–1606, 1950.

[49] R. M. Harrison, F. Romano, T. E. Ouldridge, A. A. Louis, and J. P. Doye, "Coarse-grained modelling of strong DNA bending ii: Cyclization," *arXiv preprint arXiv:1506.09008*, 2015.

[50] Q. Du, A. Kotlyar, and A. Vologodskii, "Kinking the double helix by bending deformation," *Nucleic acids research*, vol. 36, no. 4, pp. 1120–1128, 2008.

[51] A. P. Fields, E. A. Meyer, and A. E. Cohen, "Euler buckling and nonlinear kinking of double-stranded dna," *Nucleic acids research*, vol. 41, no. 21, pp. 9881–9890, 2013.

[52] W. Hwang, "Calculation of conformation-dependent biomolecular forces," *The Journal of chemical physics*, vol. 127, no. 17, p. 175 104, 2007.

[53] G. I. Bell, "Models for the specific adhesion of cells to cells," *Science*, vol. 200, no. 4342, pp. 618–627, 1978.

[54] T. T. Le and H. D. Kim, "Measuring shape-dependent looping probability of DNA," *Biophysical Journal*, vol. 104, no. 9, pp. 2068–2076, 2013.

[55] J. Jeong, T. T. Le, and H. D. Kim, "Single-molecule fluorescence studies on DNA looping," *Methods*, vol. 105, pp. 34–43, 2016.

[56] A. Vologodskii, *Biophysics of DNA*. Cambridge University Press, 2015.

[57] K. A. Schallhorn, K. O. Freedman, J. M. Moore, J. Lin, and P. C. Ke, "Single-molecule DNA flexibility in the presence of base-pair mismatch," *Applied Physics Letters*, vol. 87, no. 3, p. 033 901, 2005.

[58] T. Förster, "Zwischenmolekulare energiewanderung und fluoreszenz," *Annalen der physik*, vol. 437, no. 1-2, pp. 55–75, 1948.

[59] M. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, "Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy," *Biophysical journal*, vol. 86, no. 4, pp. 2530–2537, 2004.

[60] M. Martin-Fernandez, C. Tynan, and S. Webb, "A 'pocket guide'to total internal reflection fluorescence," *Journal of microscopy*, vol. 252, no. 1, pp. 16–22, 2013.

[61] A. Gust, A. Zander, A. Gietl, P. Holzmeister, S. Schulz, B. Lalkens, P. Tinnefeld, and D. Grohmann, "A starting point for fluorescence-based single-molecule measurements in biomolecular research," *Molecules*, vol. 19, no. 10, pp. 15 824–15 865, 2014.

[62] J. T. Waters and H. D. Kim, "Equilibrium statistics of a surface-pinned semiflexible polymer," *Macromolecules*, vol. 46, no. 16, pp. 6659–6666, 2013.

[63] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic Acids Research*, vol. 31, pp. 3406–3415, 2003.

[64] Z. Haijun and O.-Y. Zhong-Can, "Bending and twisting elasticity: A revised marko-siggia model on DNA chirality," *Physical Review E*, vol. 58, no. 4, p. 4816, 1998.

[65] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix," *Nucleic acids research*, vol. 34, no. 2, pp. 564–574, 2006.

[66] F. Kriegel, N. Ermann, R. Forbes, D. Dulin, N. H. Dekker, and J. Lipfert, "Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers," *Nucleic acids research*, vol. 45, no. 10, pp. 5920–5929, 2017.

[67] M. J. Lane, T. Paner, I. Kashin, B. D. Faldasz, B. Li, F. J. Gallo, and A. S. Benight, "The thermodynamic advantage of DNA oligonucleotide 'stacking hybridization'reactions: Energetics of a DNA nick," *Nucleic acids research*, vol. 25, no. 3, pp. 611–616, 1997.

[68] Y.-Y. Wu, L. Bao, X. Zhang, and Z.-J. Tan, "Flexibility of short DNA helices with finite-length effect: From base pairs to tens of base pairs," *The Journal of Chemical Physics*, vol. 142, no. 12, 03B614_1, 2015.

[69] M. Zoli, "J-factors of short DNA molecules," *The Journal of Chemical Physics*, vol. 144, p. 214 104, 2016.

[70] Y. Zhang and D. M. Crothers, "Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization," *Biophysical Journal*, vol. 84, pp. 136–153, 2003.

[71] Y. Zhang, A. E. McEwen, D. M. Crothers, and S. D. Levene, "Statistical-mechanical theory of DNA looping," *Biophysical journal*, vol. 90, no. 6, pp. 1903–1912, 2006.

[72] X. Xu, B. J. R. Thio, and J. Cao, "Correlated local bending of a DNA double helix and its effect on DNA flexibility in the sub-persistence-length regime," *The Journal of Physical Chemistry Letters*, vol. 5, pp. 2868–2873, 2014.

[73] K. D. Whitley, M. J. Comstock, and Y. R. Chemla, "Elasticity of the transition state for oligonucleotide hybridization," *Nucleic Acids Research*, vol. 45, pp. 547–555, 2016.

[74] Q. Wang and B. M. Pettitt, "Sequence affects the cyclization of DNA minicircles," *The Journal of Physical Chemistry Letters*, vol. 7, pp. 1042–1046, 2016.

[75] J. Wang and G. Giaever, "Action at a distance along a dna," *Science*, vol. 240, pp. 300–304, 1988.

[76] J. M. G. Vilar and L. Saiz, "DNA looping in gene regulation: From the assembly of macromolecular complexes to the control of transcriptional noise," *Current Opinion in Genetics & Development*, vol. 15, pp. 136–144, 2005.

[77] Y. Tong and R. S. Manning, "Quantifying the impact of simple DNA parameters on the cyclization j-factor for single-basepair-addition families," *Scientific Reports*, vol. 8, 2018.

[78] M. Taranova, A. D. Hirsh, N. C. Perkins, and I. Andricioaei, "Role of microscopic flexibility in tightly curved dna," *The Journal of Physical Chemistry B*, vol. 118, pp. 11 028–11 036, 2014.

[79] D. A. Sivak and P. L. Geissler, "Consequences of local inter-strand dehybridization for large-amplitude bending fluctuations of double-stranded dna," *The Journal of Chemical Physics*, vol. 136, p. 045 102, 2012.

[80] H. Shroff, D. Sivak, J. J. Siegel, A. L. McEvoy, M. Siu, A. Spakowitz, P. L. Geissler, and J. Liphardt, "Optical measurement of mechanical forces inside short DNA loops," *Biophysical Journal*, vol. 94, pp. 2179–2186, 2008.

[81] J. Shin, O.-C. Lee, and W. Sung, "How a short double-stranded DNA bends," *The Journal of Chemical Physics*, vol. 142, p. 155 101, 2015.

[82] G. Rosanio, J. Widom, and O. C. Uhlenbeck, "In vitro selection of dnas with an increased propensity to form small circles," *Biopolymers*, vol. 103, pp. 303–320, 2015.

[83] K. Rippe, "Making contacts on a nucleic acid polymer," *Trends in Biochemical Sciences*, vol. 26, pp. 733–740, 2001.

[84] H. Qu, Y. Wang, C.-Y. Tseng, and G. Zocchi, "Critical torque for kink formation in double-stranded dna," *Physical Review X*, vol. 1, 2011.

[85] G. Mustafa, C.-Y. Chuang, W. A. Roy, M. M. Farhath, N. Pokhrel, Y. Ma, K. Nagasawa, E. Antony, M. J. Comstock, S. Basu, and H. Balci, "A force sensor that converts fluorescence signal into force measurement utilizing short looped dna," *Biosensors and Bioelectronics*, vol. 121, pp. 34–40, 2018.

[86] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, "Mechanical separation of the complementary strands of dna," *Proceedings of the National Academy of Sciences*, vol. 94, no. 22, pp. 11 935–11 940, 1997.

[87] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, "Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins," *Proceedings of the National Academy of Sciences*, vol. 103, no. 16, pp. 6190–6195, 2006.

[88] K. Hatch, C. Danilowicz, V. Coljee, and M. Prentiss, "Demonstration that the shear force required to separate short double-stranded DNA does not increase significantly with sequence length for sequences longer than 25 base pairs," *Physical Review E*, vol. 78, no. 1, p. 011 920, 2008.

[89] S. Guo, Q. Tang, M. Yao, H. You, S. Le, H. Chen, and J. Yan, "Structural-elastic determination of the force-dependent transition rate of biomolecules," *Chemical Science*, 2018.

[90] J. F. Marko and E. D. Siggia, "Bending and twisting elasticity of dna," *Macromolecules*, vol. 27, pp. 981–988, 1994.

[91] N. Laurens, D. A. Rusling, C. Pernstich, I. Brouwer, S. E. Halford, and G. J. L. Wuite, "DNA looping by foki: The impact of twisting and bending rigidity on protein-induced looping dynamics," *Nucleic Acids Research*, vol. 40, pp. 4988–4997, 2012.

[92] P.-M. Lam and Y. Zhen, "Cyclization of short DNA fragments," *Physica A: Statistical Mechanics and its Applications*, vol. 482, pp. 569–572, 2017.

[93] H. Kashida, A. Kurihara, H. Kawai, and H. Asanuma, "Orientation-dependent fret system reveals differences in structures and flexibilities of nicked and gapped DNA duplexes," *Nucleic Acids Research*, vol. 45, e105–e105, 2017.

[94] C. Joseph, C.-Y. Tseng, G. Zocchi, and T. Tlusty, "Asymmetric effect of mechanical stress on the forward and reverse reaction catalyzed by an enzyme," *PLoS ONE*, vol. 9, C. M. Aegerter, Ed., e101442, 2014.

[95] B. Hua, K. Y. Han, R. Zhou, H. Kim, X. Shi, S. C. Abeysirigunawardena, A. Jain, D. Singh, V. Aggarwal, S. A. Woodson, and T. Ha, "An improved surface passivation method for single-molecule studies," *Nature Methods*, vol. 11, pp. 1233–1236, 2014.

[96] G. J. Gemmen, "DNA looping by two-site restriction endonucleases: Heterogeneous probability distributions for loop size and unbinding force," *Nucleic Acids Research*, vol. 34, pp. 2864–2877, 2006.

[97] S. Geggier and A. Vologodskii, "Sequence dependence of DNA bending rigidity," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 15 421–15 426, 2010.

[98] J. J. McCann, U. B. Choi, L. Zheng, K. Weninger, and M. E. Bowen, "Optimizing methods to recover absolute fret efficiency from immobilized single molecules," *Biophysical journal*, vol. 99, no. 3, pp. 961–970, 2010.

[99] T. Ha, A. Y. Ting, J. Liang, W. B. Caldwell, A. A. Deniz, D. S. Chemla, P. G. Schultz, and S. Weiss, "Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism," *Proceedings of the National Academy of Sciences*, vol. 96, no. 3, pp. 893–898, 1999.

[100] K. D. Grasser, S.-H. Teo, K.-B. Lee, R. W. Broadhurst, C. Rees, C. H. Hardman, and J. O. Thomas, "Dna-binding properties of the tandem hmg boxes of high-mobility-group protein 1 (hmg1)," *European journal of biochemistry*, vol. 253, no. 3, pp. 787–795, 1998.

[101] P. M. Pil, C. S. Chow, and S. J. Lippard, "High-mobility-group 1 protein mediates DNA bending as determined by ring closures," *Proceedings of the National Academy of Sciences*, vol. 90, no. 20, pp. 9465–9469, 1993.

[102] D. Ho, J. L. Zimmermann, F. A. Dehmelt, U. Steinbach, M. Erdmann, P. Severin, K. Falter, and H. E. Gaub, "Force-driven separation of short double-stranded dna," *Biophysical journal*, vol. 97, no. 12, pp. 3158–3167, 2009.

[103] S. M. Freier, D. Alkema, A. Sinclair, T. Neilson, and D. H. Turner, "Contributions of dangling end stacking and terminal base-pair formation to the stabilities of xggccp, xccggp, xggccyp, and xccggyp helixes," *Biochemistry*, vol. 24, pp. 4533–4539, 1985.

[104] N. Douarche and S. Cocco, "Protein-mediated DNA loops: Effects of protein bridge size and kinks," *Physical Review E*, vol. 72, 2005.

[105] J. E. Bronson, J. Fei, J. M. Hofman, R. L. Gonzalez, and C. H. Wiggins, "Learning rates and states from biophysical time series: A bayesian approach to model selection and single-molecule fret data," *Biophysical Journal*, vol. 97, pp. 3196–3205, 2009.

[106] S. Bommarito, N. Peyret, and J. S. Jr, "Thermodynamic parameters for DNA sequences with dangling ends," *Nucleic Acids Research*, vol. 28, pp. 1929–1934, 2000.

[107] C. E. Aitken, R. A. Marshall, and J. D. Puglisi, "An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments," *Biophysical Journal*, vol. 94, pp. 1826–1835, 2008.

[108] P.-G. de Gennes, "Maximum pull out force on DNA hybrids," *Comptes Rendus de l'Académie des Sciences - Series IV - Physics*, vol. 2, no. 10, pp. 1505–1508, 2001.

[109] K. D. Whitley, M. J. Comstock, and Y. R. Chemla, "Ultrashort nucleic acid duplexes exhibit long wormlike chain behavior with force-dependent edge effects," *Physical Review Letters*, vol. 120, no. 6, 2018.

[110] E. Protozanova, P. Yakovchuk, and M. D. Frank-amenetskii, "Stacked–unstacked equilibrium at the nick site of DNA," *Journal of Molecular Biology*, vol. 342, no. 3, pp. 775–785, 2004.

[111] A. Krueger, E. Protozanova, and M. D. Frank-Kamenetskii, "Sequence-dependent basepair opening in DNA double helix," *Biophysical Journal*, vol. 90, no. 9, pp. 3091–3099, 2006.

[112] A. Cournac and J. Plumbridge, "DNA looping in prokaryotes: Experimental and theoretical approaches," *Journal of bacteriology*, vol. 195, no. 6, pp. 1109–1119, 2013.

[113] N. A. Becker, T. L. Schwab, K. J. Clark, and L. J. Maher III, "Bacterial gene control by DNA looping using engineered dimeric transcription activator like effector (tale) proteins," *Nucleic acids research*, vol. 46, no. 5, pp. 2690–2696, 2018.

[114] Y. Yan, F. Leng, L. Finzi, and D. Dunlap, "Protein-mediated looping of DNA under tension requires supercoiling," *Nucleic acids research*, vol. 46, no. 5, pp. 2370–2379, 2018.

[115] L. D. Brennan, R. A. Forties, S. S. Patel, and M. D. Wang, "DNA looping mediates nucleosome transfer," *Nature communications*, vol. 7, p. 13 337, 2016.

[116] M. S. van Ruiten and B. D. Rowland, "SMC complexes: Universal DNA looping machines with distinct regulators," *Trends in Genetics*, 2018.

[117] C. Tardin, "The mechanics of DNA loops bridged by proteins unveiled by single-molecule experiments," *Biochimie*, vol. 142, pp. 80–92, 2017.

[118] Y. Shang, N. Zhang, P. Zhu, Y. Luo, K. Huang, W. Tian, and W. Xu, "Restriction enzyme cutting site distribution regularity for DNA looping technology," *Gene*, vol. 534, no. 2, pp. 222–228, 2014.

[119] S. M. Giovan, A. Hanke, and S. D. Levene, "DNA cyclization and looping in the wormlike limit: Normal modes and the validity of the harmonic approximation," *Biopolymers*, vol. 103, no. 9, pp. 528–538, 2015.

[120] B. Joffroy, Y. O. Uca, D. Prešern, J. P. K. Doye, and T. L. Schmidt, "Rolling circle amplification shows a sinusoidal template length-dependent amplification bias," *Nucleic acids research*, vol. 46, no. 2, pp. 538–545, 2017.

[121] K. Olsen and J. Bohr, "Geometry of the toroidal n-helix: Optimal-packing and zero-twist," *New Journal of Physics*, vol. 14, no. 2, p. 023 063, 2012.

[122] S. Panyukov and Y. Rabin, "Fluctuating elastic rings: Statics and dynamics," *Physical Review E*, vol. 64, no. 1, p. 011 909, 2001.

[123] T. Sutthibutpong, C. Matek, C. Benham, G. G. Slade, A. Noy, C. Laughton, J. P. K. Doye, A. A. Louis, and S. A. Harris, "Long-range correlations in the mechanics of small DNA circles under topological stress revealed by multi-scale simulation," *Nucleic acids research*, vol. 44, no. 19, pp. 9121–9130, 2016.

[124] A. R. Haeusler, K. A. Goodson, T. D. Lillian, X. Wang, S. Goyal, N. C. Perkins, and J. D. Kahn, "Fret studies of a landscape of lac repressor-mediated DNA loops," *Nucleic acids research*, vol. 40, no. 10, pp. 4432–4445, 2012.

[125] M. Ganji, I. A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C. H. Haering, and C. Dekker, "Real-time imaging of DNA loop extrusion by condensin," *Science*, vol. 360, pp. 102–105, 2018.

[126] J. F. Marko, P. De Los Rios, A. Barducci, and S. Gruber, "Dna-segment-capture model for loop extrusion by structural maintenance of chromosome (SMC) protein complexes," *bioRxiv*, 2018.

[127] V. Vogel and M. Sheetz, "Local force and geometry sensing regulate cell functions," *Nature reviews Molecular cell biology*, vol. 7, no. 4, p. 265, 2006.

[128] X. Zheng and A. Vologodskii, "Theoretical analysis of disruptions in DNA minicircles," *Biophysical Journal*, vol. 96, pp. 1341–1349, Feb. 2009.

[129] J. Zhang, Y. Yan, S. Samai, and D. S. Ginger, "Dynamic melting properties of photoswitch-modified dna: Shearing versus unzipping," *The Journal of Physical Chemistry B*, vol. 120, pp. 10 706–10 713, Oct. 2016.

[130] C. Yuan, E. Rhoades, X. W. Lou, and L. A. Archer, "Spontaneous sharp bending of dna: Role of melting bubbles," *Nucleic Acids Research*, vol. 34, pp. 4554–4560, Sep. 2006.

[131] K.-C. Tham, R. Kanaar, and J. H. G. Lebbink, "Mismatch repair and homeologous recombination," *DNA Repair*, vol. 38, pp. 75–83, Feb. 2016.

[132] S. R. Tee and Z. Wang, "How well can DNA rupture dna? shearing and unzipping forces inside DNA nanostructures," *ACS Omega*, vol. 3, pp. 292–301, Jan. 2018.

[133] M. Sharma, A. V. Predeus, S. Mukherjee, and M. Feig, "DNA bending propensity in the presence of base mismatches: Implications for DNA repair," *The Journal of Physical Chemistry B*, vol. 117, pp. 6194–6205, 2013.

[134] J. Brown, T. Brown, *et al.*, "Affinity of mismatch-binding protein muts for heteroduplexes containing different mismatches," *Biochemical Journal*, vol. 354, no. 3, pp. 627–633, 2001.

[135] D. L. Floyd, S. C. Harrison, and A. M. Van Oijen, "Analysis of kinetic intermediates in single-particle dwell-time distributions," *Biophysical journal*, vol. 99, no. 2, pp. 360–366, 2010.

[136] G. Bel, B. Munsky, and I. Nemenman, "The simplicity of completion time distributions for common complex biochemical processes," *Physical biology*, vol. 7, no. 1, p. 016 003, 2009.

[137] J. SantaLucia and D. Hicks, "The thermodynamics of DNA structural motifs," *Annual Review of Biophysics and Biomolecular Structure*, vol. 33, pp. 415–440, Jun. 2004.

[138] G. Rossetti, P. D. Dans, I. Gomez-Pinto, I. Ivani, C. Gonzalez, and M. Orozco, "The structural impact of DNA mismatches," *Nucleic Acids Research*, vol. 43, pp. 4309–4321, Mar. 2015.

[139] M. Mosayebi, A. A. Louis, J. P. K. Doye, and T. E. Ouldridge, "Force-induced rupture of a DNA duplex: From fundamentals to force sensors," *ACS Nano*, vol. 9, pp. 11 993–12 003, Nov. 2015.

[140] T. A. Lionberger, D. Demurtas, G. Witz, J. Dorier, T. Lillian, E. Meyhöfer, and A. Stasiak, "Cooperative kinking at distant sites in mechanically stressed dna," *Nucleic Acids Research*, vol. 39, pp. 9820–9832, Sep. 2011.

[141] T. A. Kunkel and D. A. Erie, "Eukaryotic mismatch repair in relation to DNA replication," *Annual Review of Genetics*, vol. 49, pp. 291–313, Nov. 2015.

[142] J. D. Kahn, E. Yun, and D. M. Crothers, "Detection of localized DNA flexibility," *Nature*, vol. 368, pp. 163–166, Mar. 1994.

[143] L. J. Friedman, J. Chung, and J. Gelles, "Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence," *Biophysical journal*, vol. 91, no. 3, pp. 1023–1031, 2006.

[144] J. Jeong and H. D. Kim, "Determinants of cyclization-decyclization kinetics of short DNA with sticky ends," *bioRxiv*, 2018.

[145] A. Granzhan, N. Kotera, and M.-P. Teulade-Fichou, "Finding needles in a base-stack: Recognition of mismatched base pairs in DNA by small molecules," *Chemical Society Reviews*, vol. 43, p. 3630, 2014.

[146] A. Dittmore, S. Brahmachari, Y. Takagi, J. F. Marko, and K. C. Neuman, "Supercoiling DNA locates mismatches," *Physical review letters*, vol. 119, no. 14, p. 147 801, 2017.

[147] N. Chatterjee and G. C. Walker, "Mechanisms of DNA damage, repair, and mutagenesis," *Environmental and Molecular Mutagenesis*, vol. 58, pp. 235–263, 2017.

[148] S. Chakraborty, P. J. Steinbach, D. Paul, H. Mu, S. Broyde, J.-H. Min, and A. Ansari, "Enhanced spontaneous DNA twisting/bending fluctuations unveiled by fluorescence lifetime distributions promote mismatch recognition by the rad4 nucleotide excision repair complex," *Nucleic Acids Research*, vol. 46, pp. 1240–1255, Dec. 2017.

[149] J. Cadet and J. R. Wagner, "DNA base damage by reactive oxygen species, oxidizing agents, and uv radiation," *Cold Spring Harbor Perspectives in Biology*, vol. 5, a012559–a012559, Feb. 2013.

[150] J. T. Waters and H. D. Kim, "Force distribution in a semiflexible loop," *Physical Review E*, vol. 93, no. 4, p. 043 315, 2016.

[151] I. I. Cisse, H. Kim, and T. Ha, "A rule of seven in watson-crick base-pairing of mismatched sequences," *Nature structural & molecular biology*, vol. 19, no. 6, p. 623, 2012.

[152] H. A. Erlich, D. Gelfand, and J. J. Sninsky, "Recent advances in the polymerase chain reaction," *Science*, vol. 252, no. 5013, pp. 1643–1651, 1991.

[153] A. Raj, P. Van Den Bogaard, S. A. Rifkin, A. Van Oudenaarden, and S. Tyagi, "Imaging individual mrna molecules using multiple singly labeled probes," *Nature methods*, vol. 5, no. 10, p. 877, 2008.

[154] G. Seelig, D. Soloveichik, D. Y. Zhang, and E. Winfree, "Enzyme-free nucleic acid logic circuits," *science*, vol. 314, no. 5805, pp. 1585–1588, 2006.

[155] X. Wang and T. Ha, "Defining single molecular forces required to activate integrin and notch signaling," *Science*, vol. 340, no. 6135, pp. 991–994, 2013.

[156] N. C. Seeman, "Nanomaterials based on dna," *Annual review of biochemistry*, vol. 79, pp. 65–87, 2010.

[157] I. Donmez and S. S. Patel, "Mechanisms of a ring shaped helicase," *Nucleic acids research*, vol. 34, no. 15, pp. 4216–4224, 2006.

[158] F. Vella, "Molecular biology of the cell: By b alberts, d bray, j lewis, m raff, k roberts and jd watson. pp 1361. garland publishing, new york and london. 1994," *Biochemical Education*, vol. 22, no. 3, pp. 164–164, 1994.

[159] M. J. Comstock, T. Ha, and Y. R. Chemla, "Ultrahigh-resolution optical trap with single-fluorophore sensitivity," *Nature methods*, vol. 8, no. 4, p. 335, 2011.

[160] G. Zocchi, "Controlling proteins through molecular springs," *Annual review of biophysics*, vol. 38, pp. 75–88, 2009.

[161] A. Grove, A. Galeone, L. Mayol, and P. E. Geiduschek, *Localized DNA flexibility contributes to target site selection by dna-bending proteins*, 1996.

[162] ——, "On the connection between inherent DNA flexure and preferred binding of hydroxymethyluracil-containing DNA by the type ii dna-binding protein tf1," *Journal of molecular biology*, vol. 260, no. 2, pp. 196–206, 1996.

[163] A. Ansari and S. V. Kuznetsov, "Dynamics and mechanism of dna-bending proteins in binding site recognition," in *Biophysics of DNA-Protein Interactions*, Springer, 2010, pp. 107–142.

[164] R. J. Isaacs and H. P. Spielmann, "A model for initial DNA lesion recognition by ner and mmr based on local conformational flexibility," *DNA repair*, vol. 3, no. 5, pp. 455–464, 2004.

[165] G. Neuert, C. H. Albrecht, and H. E. Gaub, "Predicting the rupture probabilities of molecular bonds in series," *Biophysical journal*, vol. 93, no. 4, pp. 1215–1223, 2007.